

Université de Montréal

**Cohérentisme historique et pratiques classificatoires : la
phylogénétique contemporaine comme cas d'étude**

Par

François Papale

Département de philosophie, Faculté des études supérieures et postdoctorales

Thèse présentée en vue de l'obtention du grade de Ph.D. en philosophie

Décembre 2021

© François Papale, 2021

Université de Montréal

Département de philosophie, Faculté des arts et des sciences

Cette thèse intitulée

**Cohérentisme historique et pratiques classificatoires : la phylogénétique
contemporaine comme cas d'étude**

Présentée par

François Papale

A été évaluée par un jury composé des personnes suivantes

Molly Kao

Présidente-rapporteuse

Frédéric Bouchard

Directeur de recherche

Éric Bapteste (Sorbonne Université)

Codirecteur

François Duchesneau

Membre du jury

Pierre-Olivier Méthot (Université Laval)

Examineur externe

Christopher Cameron (Département des sciences biologiques)

Représentant du doyen

Résumé

L'objectif de cette thèse est de fournir un cadre théorique pour l'analyse philosophique des pratiques classificatoires en sciences. La classification, définie comme étant un geste épistémique dont l'objectif est de fournir une description du monde par découpage et discrimination, et par mise en relation des catégories ainsi formées, est un point de rencontre naturel entre philosophie et sciences. La tradition des espèces naturelles, considérée en philosophie des sciences comme la principale approche disponible pour l'analyse des pratiques classificatoires, est toutefois lacunaire. Dans le cadre de cette thèse, je démontre les faiblesses de cette approche et le besoin de mobiliser d'autres ressources épistémologiques pour penser la classification. À cet effet, je m'inspire de travaux en philosophie de la mesure s'inscrivant dans un courant de pensée que j'appelle le cohérentisme historique. Ces approches mettent l'accent sur les pratiques inférentielles et sur la charge théorique des observations en plus d'insister sur le dynamisme scientifique et la multitude de facteurs qui l'influencent (principe de respect, impératif de progrès, développements théoriques, nouveaux outils [mathématiques ou matériels], nouvelles données, histoire du champ disciplinaire, structure sociale de la recherche, etc.).

Une fois ce cadre d'analyse épistémologique présenté en détail, je le mobilise pour étudier un ensemble de pratiques classificatoires en sciences, soit les pratiques phylogénétiques contemporaines. La phylogénétique est la discipline biologique qui vise à retracer les liens généalogiques qui unissent les êtres vivants. Les gènes et organismes sont ainsi regroupés au sein d'unités taxonomiques qui sont par la suite mises en relation. Ce champ disciplinaire a longtemps été associé de façon intime et exclusive au modèle arborescent, lui-même associé à la théorie de l'évolution par voie de sélection naturelle : les liens phylogénétiques entre organismes sont généralement représentés sous la forme d'un arbre, soit l'Arbre du vivant. Or, les dernières décennies ont vu émerger, au sein de la discipline, un compétiteur : le modèle réticulé. En analysant les différences et similarités qui lient ces deux types de modèles (modèles arborescents

et modèles réseaux), je démontre que la transition d'un modèle à l'autre n'influence que superficiellement la pratique puisque la majorité de ses déterminants demeure inchangée.

En bref, cette thèse présente un cadre d'analyse philosophique pour appréhender les pratiques classificatoires en sciences. J'importe des principes et des outils issus de travaux de la philosophie de la mesure afin de développer une philosophie de la classification qui comble les lacunes de la tradition des espèces naturelles. En s'ancrant dans ce que j'appelle le cohérentisme historique, cette thèse contribue non seulement à la bonne compréhension des pratiques classificatoires d'un point de vue épistémologique, mais aussi à la bonne compréhension des pratiques phylogénétiques contemporaines et de l'importante transition qui éloigne cette discipline de son modèle classique, soit l'Arbre du vivant.

Mots-clés : épistémologie, philosophie des sciences, philosophie de la biologie, cohérentisme historique, classification, espèces naturelles, éliminativisme, phylogénétique, Arbre du vivant, réseaux.

Abstract

The main objective of this dissertation is to offer a philosophical framework for the analysis of scientific classificatory practices. Classification, here defined as an epistemic action that describes the world by breaking down phenomena into categories and by establishing relevant relationships between these categories, is a natural bridge between philosophy and science. The tradition of natural kinds, which is currently the main approach in philosophy of science for analysing classificatory practices, has important shortcomings when it comes to epistemologically analyzing scientific classification. In this dissertation, I highlight these weaknesses and the consequent need to develop a novel framework for tackling classificatory practices. To achieve this, I draw on an existing trend in philosophy of measurement, which I coined historical coherentism. This approach is centered on the analysis of inferential practices and stresses the importance of the theoretical charge of observations. It also focuses on the dynamic nature of scientific fields and on the variety of factors that determine scientific progress (principle of respect, imperative of progress, theoretical developments, mathematical and material innovations, new data, the history of a field of research, the social structure of the scientific community, etc.).

After the detailed presentation and exploration of this epistemological framework, I use it to tackle a set of scientific classificatory practices, namely contemporary phylogenetic practices. Phylogenetics is the biological discipline that aims to reconstruct the genealogical relationships uniting living beings. Genes and organisms are grouped into operative taxonomic units, and these units are then connected within a system. Since its inception, phylogenetics has been intimately tied to tree-based models, with these tree-based models themselves tied to the theory of evolution by means of natural selection: the phylogenetic system connecting all living organisms takes the form of a tree, the Tree of Life. In the last few decades, however, an alternative type of models has gained more attention and support within the field: network-based models. Using historical coherentism, I analyze the differences and similarities between the two competing types of models (tree-based models and network-based models) and show that the transition

from one to the other only superficially changes phylogenetic practices. Indeed, many factors that constrain these practices remain unchanged whether researchers use tree-based or network-based models.

In short, this thesis presents a philosophical framework for the analysis of classificatory practices in sciences. I use various principles and tools imported from the philosophy of measurement to develop a philosophy of classification that covers the weaknesses of the tradition of natural kinds. By anchoring my work in historical coherentism, I contribute not only to our understanding of classificatory practices, from an epistemological perspective, but also to a better understanding of contemporary phylogenetic practices and of the significant transition that leads this discipline away from its classical model, the Tree of Life.

Keywords: epistemology, philosophy of sciences, philosophy of biology, historical coherentism, classification, natural kinds, eliminativism, phylogenetics, Tree of Life, networks.

Table des matières

Résumé.....	5
Abstract.....	7
Table des matières.....	9
Liste des tableaux.....	15
Liste des figures.....	17
Remerciements.....	19
Introduction de la thèse.....	23
Chapitre 1 – Le problème de la coordination et la structure épistémique de la phylogénétique.....	47
1.1 Introduction.....	47
1.2 Le problème de la coordination.....	48
1.3 Généralisation du problème de la coordination.....	51
1.4 Le problème de la coordination et les pratiques classificatoires : l'exemple de la cladistique hennigienne.....	53
1.4.1 Le fonctionnement inférentiel de la cladistique hennigienne.....	53
1.5 Le rapport de justification mutuelle unissant la cladistique hennigienne et la théorie l'évolution par voie de sélection naturelle.....	57
1.5.1 La structure épistémique de la cladistique hennigienne et le problème de la coordination.....	62
1.5.2 Un autre aspect du problème de la coordination : la formation de catégories.....	68
1.6 La structure justificative de la biologie comparative est dynamique, comme illustré par la notion d'homologie.....	70
1.6.1 L'homologie au sein de la structure épistémique de la phylogénétique.....	72

1.6.2	L'homologie prédarwinienne.....	77
1.6.3	L'homologie et la mise en place de la phylogénétique	80
1.6.4	La monophylie comme cible de la phylogénétique et critère définitionnel de l'homologie.....	84
1.6.5	Cohérentisme historique et hélice phylogénétique	88
1.7	Conclusion.....	92
Chapitre 2 – Vers une philosophie de la classification inspirée du problème de la coordination et de la philosophie de la mesure		
		95
2.1	Introduction	95
2.2	Les réponses philosophiques au problème de la coordination (20 ^e siècle)	96
2.2.1	Le conventionnalisme	96
2.2.2	Opérationnalisme	104
2.2.3	Le constructivisme et le cohérentisme	107
2.3	Le cohérentisme historique	112
2.3.1	Hasok Chang : itérations épistémiques et cohérentisme progressif.....	113
2.3.2	Bastian van Fraassen.....	123
2.3.2.1	La vision synoptique.....	123
2.3.2.2	Van Fraassen et l'approche sémantique des théories scientifiques	128
2.3.3	Approches centrées sur les modèles	129
2.3.4	Une critique des approches cohérentistes	135
2.4	Conclusion.....	145
Chapitre 3 – L'arche des espèces naturelles.....		
		151
3.1	Introduction	151
3.2	Les fondements historiques du complexe discursif.....	153

3.2.1	L'Arbre de Porphyre.....	153
3.2.2	Les essences lockéennes.....	154
3.2.3	La cristallisation du schème de pensée relié aux espèces naturelles.....	156
3.3	Les théories contemporaines des espèces naturelles	160
3.3.1	Les approches essentialistes.....	161
3.3.1.1	L'essence intrinsèque.....	164
3.3.1.2	L'essence microstructurale.....	169
3.3.1.3	L'essentialisme, en bref	171
3.3.2	Les approches pluralistes.....	172
3.3.2.1	Boyd, la thèse de l'accommodement et les groupes homéostatiques de propriétés	172
3.3.2.2	Dupré et le réalisme dévergondé	177
3.3.2.3	Les théories dites pluralistes, en bref.....	180
3.3.3	Les approches non pluralistes, non essentialistes.....	181
3.3.3.1	Khalidi et les espèces naturelles en tant que nœuds au sein de réseaux causaux	182
3.3.3.2	Slater et les groupes stables de propriétés	185
3.3.3.3	Les approches non pluralistes non essentialistes, en bref	187
3.4	Conclusion.....	188
Chapitre 4 – Éliminativisme et espèces naturelles		191
4.1	Introduction	191
4.2	L'éliminativisme et les espèces naturelles.....	194
4.2.1	Deux arguments éliminativistes.....	194
4.2.1.1	Un cadre d'analyse multidimensionnel pour appréhender les pratiques de classification non arbitraires	195

4.2.1.2	La naturalité des espèces naturelles et l'influence des valeurs en sciences ...	198
4.2.1.3	Les approches éliminativistes de Brigandt et de Ludwig, en bref, ainsi qu'une première esquisse de la mienne	201
4.2.2	Un argument contre l'éliminativisme et ma réponse à cet argument	203
4.3	De l'utilité épistémologique de la notion d'espèce naturelle.....	210
4.3.1	L'homologie phylogénétique et les groupes de propriétés homéostatiques.....	210
4.3.2	Brigandt et l'utilité épistémologique de la notion d'espèce naturelle	212
4.3.3	Deux arguments auxiliaires au rejet de l'utilité épistémologique de la notion d'espèce naturelle.....	219
4.4	Conclusion.....	224
Chapitre 5 – L'homologie et la transition moléculaire.....		227
5.1	Introduction	227
5.2	La transition moléculaire	228
5.3	L'homologie moléculaire.....	232
5.4	De l'analyse épistémologique de l'homologie en termes d'espèce naturelle.....	236
5.5	Les homologies moléculaire et morphologique analysées à l'aune du cohérentisme historique	240
5.6	Conclusion.....	246
Chapitre 6 – Analyse épistémologique de la pensée arborescente et de la phylogénétique traditionnelle.....		249
6.1	Introduction	249
6.2	La phylogénétique traditionnelle : une itération épistémique de la systématique	250
6.2.1	La pensée arborescente : le cadre théorique de la phylogénétique traditionnelle.....	251
6.2.2	La pensée arborescente et le tournant génomique	257

6.2.2.1	Méthodes inférentielles de la phylogénétique traditionnelle.....	259
6.2.2.2	Méthodes inférentielles de la phylogénomique arborescente	262
6.2.2.3	Phylogénomique et stabilité de la pensée arborescente	266
6.3	Une première analyse critique : trois constats	268
6.3.1	Premier constat : la pensée arborescente implique une série d’amalgames	269
6.3.1.1	Premier amalgame : modèle arborescent en opposition aux pensées du groupe et du développement.....	270
6.3.1.2	Second amalgame : pratiques phylogénétiques et modèle arborescent.....	272
6.3.1.3	Troisième amalgame : phylogénétique et évolution par voie de sélection naturelle.....	274
6.3.1.4	Un bref retour sur le premier constat	278
6.3.2	Deuxième constat : le modèle arborescent est un principe ontologique	279
6.3.3	Troisième constat : l’ancrage philosophique de la phylogénétique est fondationnaliste et donc obsolète	283
6.3.4	Un bref retour sur cette première analyse critique	286
6.4	Le placement phylogénétique de <i>Methanonatronarchaeia</i>	288
6.4.1	Le débat.....	288
6.4.2	Analyse critique du débat	293
6.4.3	Un bref retour sur le débat concernant <i>Methanonatronarchaeia</i>	298
6.5	Conclusion.....	299
Chapitre 7 – De la comparaison entre deux modèles phylogénétiques : les réseaux et les arbres.....		303
7.1	Introduction	303
7.2	Qu’est-ce qu’un réseau phylogénétique et en quoi cela diffère d’un arbre phylogénétique?.....	305

7.2.1	Les réseaux phylogénétiques et la théorie des graphes.....	305
7.2.1.1	Les réseaux abstraits.....	310
7.2.1.2	Les réseaux explicites.....	313
7.2.2	Une brève histoire des réseaux, en phylogénétique et ailleurs	318
7.2.3	Une première analyse comparative des itérations contemporaines de la phylogénétique : les approches réseaux et la pensée arborescente	323
7.3	Les débats entourant les réseaux phylogénétiques	330
7.3.1	De la compatibilité entre les réseaux phylogénétiques et le cadre darwinien	331
7.3.2	Réalisme et modèles phylogénétiques	336
7.3.3	L'utilité respective des réseaux et des arbres	339
7.3.4	De la place des lignées (et conséquemment de la phylogénétique) en biologie de l'évolution	343
7.3.5	La pensée populationnelle, la phylogénétique et le traitement de l'information biologique	348
7.4	Conclusion.....	356
	Conclusion de la thèse.....	361
	Références bibliographiques.....	385

Liste des tableaux

Tableau 7.1 – Synthèse de l'analyse comparative de la phylogénétique traditionnelle et de la phylogénétique centrée sur les réseaux	325
---	-----

Liste des figures

Figure 1.1 – Deux hypothèses phylogénétiques pour le groupe A (<i>Aves</i>), B (<i>Crocodylia</i>) et C (les autres reptiles)	56
Figure 1.2 – Diagramme de la divergence des taxons selon Darwin	61
Figure 1.3 – Exemples d'hypothèses phylogénétiques variées pour le groupe A, B, C avec des contraintes théoriques relâchées	64
Figure 1.4 – Deux représentations de la structure justificative des pratiques scientifiques.....	90
Figure 3.1 – L'arche des espèces naturelles.....	159
Figure 3.2 – Les principales approches contemporaines constitutives de la tradition des espèces naturelles	161
Figure 3.3 – Un arbre phylogénétique formé de trois espèces	166
Figure 4.1 – Groupes de propriétés et inférences inductives.....	216
Figure 5.1 – Un exemple d'alignement de séquences génétiques	244
Figure 5.2 – Deux alignements possibles pour les séquences 1 et 2	245
Figure 6.1 – Évolution par voie de sélection naturelle avec et sans formation de lignées	278
Figure 7.1 – Un exemple de réseau phylogénétique enraciné mais abstrait	309
Figure 7.2 – Un réseau de bipartitions et les deux arbres qu'il représente	311

À toutes ces personnes qui se font catégoriser et qui en souffrent

Remerciements

Je veux d'abord remercier Cassandra Ville, ma partenaire de vie, qui donne un sens aux choses que je fais ainsi qu'à celles que je ne fais pas. Tu m'as poussé, encouragé et même forcé à bien profiter de la vie de thésard. Merci pour tout, Cass. Au plaisir de te rendre la pareille pendant les quatre ou cinq prochaines années... Bon début de doctorat!

Je souhaite aussi expliciter ma gratitude envers mes parents, Antimo Papale et Diane David, bien que les mots les plus sophistiqués et nombreux soient inaptes à rendre honneur à tout ce qui explique ma reconnaissance envers eux. Merci de m'avoir bien outillé pour entreprendre des aventures un peu folles (comme la rédaction d'une thèse en philosophie) et de m'avoir accompagné tout au cours de mon cheminement, aussi tortueux qu'il ait pu sembler.

Je souhaite témoigner ma reconnaissance sans bornes à mes deux directeurs de recherche, Frédéric Bouchard et Éric Bapteste. Merci d'avoir toléré, entre autres choses, mes expérimentations théoriques et de considérer mes convictions philosophiques plus ou moins bien justifiées comme étant des forces. Merci de votre soutien inébranlable et de votre bonne humeur. Merci du cadre de recherche que vous m'avez offert. Merci de faire de la philo.

Je voudrais aussi remercier Louis-Patrick Haraoui de m'avoir rappelé, à un moment fort opportun, que j'aime faire de la recherche et de me motiver à continuer. Dans la même veine, j'aimerais remercier Ford Doolittle de m'avoir offert une bonne raison de finir ma thèse ainsi que d'être une source d'inspiration pour moi et une armée de jeunes personnes passionnées par l'évolution.

Un gros merci à l'ensemble des membres actuel·le·s et passé·e·s du LEIPS (Laboratoire étudiant interuniversitaire en philosophie des sciences) qui sont des allié·e·s de poids depuis un lustre et demi. Vous avez rendu la philosophie des sciences un domaine de recherche plus qu'agréable. Une mention spéciale est due à David Montminy qui est devenu, avec les années, un grand ami et un collaborateur de haute qualité. Des mentions spéciales sont aussi dues à Kevin

et Anne-Marie, Kaiser et Gagné-Julien, respectivement, pour de bons moments et de belles discussions.

Des remerciements s'imposent aussi à l'égard des membres, actuels ou non, du laboratoire AIRE (Adaptation, intégration, réticulation, évolution), codirigé par Éric Bapteste et Philippe Lopez. J'aimerais tout particulièrement remercier Charles Bernard, Philippe Lopez, Guillaume Bernard, Eduardo Corel, Jérôme Teulière et Romain Lannes pour leur convivialité et les bons moments passés ensemble, mais aussi pour leurs explications patientes et fort édifiantes.

Sur le plan institutionnel, la foule de mes alliés est trop grande pour faire l'objet d'une analyse détaillée et de remerciements individuels. Cependant, je tiens à souligner l'immensité de ma dette à l'égard du personnel (tous postes confondus) du Département de philosophie de l'Université de Montréal. En faisant vivre le département, vous faites émerger de belles possibilités. Un gros merci à l'équipe de La Plume pour son aide dans la dernière ligne droite.

Sur le plan personnel, je tiens à remercier François Léger-Bélanger pour les aurores boréales, Maxime Cléroux-Charbonneau pour la musique, Mitch (Michel Iacono) pour les rapports d'impôt, Xuan-Vink Pham pour la métaphore architecturale, Hansko pour les nombreuses heures de sommeil scandinave et toute une bande de personnes dont je recherche constamment la compagnie inmanquablement plaisante, garante de détente. Merci à Laurent Pineault d'être une chambre d'écho foucauldienne de la plus grande qualité et d'avoir donné l'impulsion intellectuelle nécessaire à la réalisation de la présente thèse.

Sur le plan financier, je tiens à remercier les institutions publiques, et du même coup l'ensemble des payeurs et des payeuses de taxes qui ont rendu mon travail possible : le Conseil de recherche en sciences humaines du Canada (CRSH), les Fonds de recherche du Québec – société et culture (FRQSC), le Département de philosophie de l'Université de Montréal, la Faculté des arts et des sciences de l'Université de Montréal et le Centre interuniversitaire de recherche sur la science et la technologie.

Je profite aussi de l'occasion pour souligner que la majorité de mes travaux ont été et sont encore effectués sur des territoires autochtones (innus et mohawks) qui n'ont jamais été cédés par ces peuples. Je souhaite ainsi rendre hommage à ces peuples et à tous les autres qui ont

occupé le magnifique territoire qu'on appelle aujourd'hui le Québec bien avant que mes ancêtres ne connaissent son existence. J'espère que la présente déclaration contribue à rendre visibles, à reconnaître et à réparer les injustices d'hier et d'aujourd'hui.

Introduction de la thèse

L'objectif de la présente thèse est de formuler une approche philosophique pour comprendre la place qu'occupent les descriptions classificatoires au sein des pratiques scientifiques, et plus largement, au sein des discours qui aspirent à être des discours de connaissances. Pour ce faire, je mobilise les outils d'un courant de pensée issu de la philosophie de la mesure que j'appelle le cohérentisme historique (Chang 2004; Mari 2005; Morrison 1998; Tal 2016; van Fraassen 2008). L'importation des outils du cohérentisme historique vers une philosophie de la classification me permet notamment de défendre une posture éliminativiste à l'égard du concept d'espèce naturelle. Ce concept est présentement au cœur des analyses des pratiques classificatoires opérées en philosophie des sciences. Or, les lacunes des théories des espèces naturelles identifiées dans cette thèse ainsi que par d'autres auteurs (Brigandt 2020; Hacking 2007a; Ludwig 2018) incitent à développer de nouvelles approches pour l'analyse épistémologique de la classification. L'utilisation du cohérentisme historique pour l'analyse des pratiques phylogénétiques contemporaines me sert à illustrer la portée explicative et analytique de cette approche, en contraste avec la tradition des espèces naturelles, en plus d'offrir d'importantes contributions en philosophie de la biologie.

Comme l'indique son nom, le cohérentisme historique fait partie de la grande famille des philosophies dites cohérentistes (Chang 2007). Celles-ci considèrent que la connaissance est le fruit de systèmes d'interactions sociales et matérielles complexes qui déterminent la teneur des connaissances et les mécanismes qui permettent de les justifier. Au sein de tels systèmes, la démonstration empirique de la validité des hypothèses, des modèles et des théories occupe une place importante, mais il ne s'agit que d'un facteur déterminant parmi d'autres. La structure qui confère aux connaissances leur légitimité est aussi déterminée par le poids de la tradition, certaines contingences matérielles (comme le développement de nouveaux outils, l'augmentation de la puissance calculatoire, etc.), les échanges entre divers champs disciplinaires (l'importation d'outils mathématiques en biologie, par exemple), la structure sociale caractérisant la communauté scientifique, les modalités de l'inscription de la communauté scientifique au sein

de la société, etc. Le cohérentisme regroupe les approches qui reconnaissent qu'une pluralité de facteurs est nécessaire pour comprendre comment les connaissances sont construites. Cette famille d'approches épistémologiques est généralement pensée en opposition au fondationnalisme, qui considère que l'empirie est la source privilégiée de la légitimité des connaissances scientifiques, leur fondation.

La spécificité du cohérentisme *historique* est de mettre l'historicité des processus de construction de connaissances à l'avant-plan des études épistémologiques. Concrètement, il s'agit de considérer que les pratiques scientifiques forment des successions d'itérations épistémiques (Chang 2004) au sein desquelles chaque itération s'appuie sur la précédente pour justifier et légitimer ses méthodes et croyances théoriques. L'analyse épistémologique des pratiques scientifiques consiste donc à étudier la structure dynamique de ces pratiques, une tâche que le cohérentisme historique invite à réaliser en adoptant la logique propre aux itérations épistémiques étudiées, mais en croisant cette logique avec un regard historique et critique sur cette même logique (van Fraassen 2008). À ma connaissance, je suis le premier à regrouper les travaux de philosophie de la mesure qui font l'objet de mon analyse au sein d'un courant unitaire (bien qu'hétérogène) que j'appelle le cohérentisme historique. Cette appellation est fortement inspirée de celle que Chang utilise pour décrire son approche, soit le cohérentisme progressiste (Chang 2007). J'ai cru bon d'utiliser un terme différent pour marquer les points similaires liant les travaux de Chang à ceux d'autres auteurs et autrices sans réduire l'ensemble à son approche.

Le développement de ressources pour appréhender les pratiques classificatoires en sciences est d'une grande importance puisque la philosophie des sciences présente actuellement des lacunes sur ce plan (Wilkins et Ebach 2014). En effet, le courant de pensée principal s'intéressant à la classification, au sein de cette sous-discipline de l'épistémologie, est la tradition des espèces naturelles (Kendig 2016). Ce courant de pensée hétérogène s'est cristallisé au 19^e siècle, mais a pris sa source aux origines de la philosophie occidentale, notamment dans les réflexions aristotéliennes sur les notions d'essence et d'espèce. En plus de témoigner de l'importance philosophique de la classification et de la tradition des espèces naturelles, ce constat souligne que les pratiques classificatoires forment une zone de contact pérenne entre la biologie et la philosophie.

Or, malgré l’ancrage profond de la tradition des espèces naturelles au sein de la philosophie occidentale et plus spécifiquement de la philosophie des sciences, je démontre dans cette thèse que cette approche est insuffisante à la réalisation d’analyses épistémologiques des structures épistémiques qui mobilisent la classification d’une façon ou d’une autre (comme mode descriptif ou comme objectif de recherche). Le cohérentisme historique offre un regard plus exhaustif sur les dynamiques qui mènent le processus scientifique vers un état des connaissances observable à un moment donné.

Les apports du cohérentisme historique sont illustrés par l’entremise d’un cas d’étude, soit les pratiques phylogénétiques contemporaines, et plus précisément la comparaison entre deux façons de pratiquer la phylogénétique : l’une centrée sur les modèles arborescents et l’autre sur les modèles réseaux. Il serait toutefois réducteur de penser que l’analyse de ces pratiques scientifiques n’est qu’un prétexte pour défendre le cohérentisme historique. La confrontation entre les modèles arborescents et les modèles réseaux qu’on observe actuellement en biologie de l’évolution promet d’être déterminante pour l’avenir ce champ d’études (DeSalle et Riley 2020; Doolittle 2000; Raoult et Koonin 2012; Velasco 2012). En effet, le modèle arborescent, et tout particulièrement son actualisation dans la croyance selon laquelle l’évolution de l’ensemble du monde vivant pourrait être représentée sous la forme d’un arbre, l’Arbre du vivant, a été un élément central de la biologie de l’évolution depuis ses débuts. Selon nombre de biologistes, le remplacement des modèles arborescents par des modèles réseaux vient remettre en question certains des fondements de la théorie de l’évolution par voie de sélection naturelle. C’est donc dire que l’opposition entre les deux types de modèles est au cœur de débats qui promettent de bouleverser le cadre théorique de la biologie de l’évolution. Il est crucial pour la philosophie de la biologie d’appréhender cette situation.

La description épistémologique des pratiques phylogénétiques, une discipline éminemment classificatoire, est un excellent point de départ. En effet, cette discipline jongle constamment avec les modèles arborescents, et depuis le tournant du 21^e siècle, avec les modèles réseaux afin d’articuler les relations entre unités taxonomiques. Autrement dit, cette discipline travaille avec des catégories d’entités biologiques qu’elle insère au sein d’un système de relations grâce aux deux types de modèles considérés. Comme mon analyse qui mobilise les

outils du cohérentisme historique le souligne, l'inscription des modèles arborescents et réseaux au sein de la structure épistémique de la phylogénétique révèle certainement des marqueurs d'opposition forte, mais aussi l'existence de points communs significatifs. Si les modèles réseaux et les modèles arborescents traduisent l'évolution des entités biologiques de façon effectivement distincte, ces représentations s'inscrivent au sein de structures inférentielles largement partagées. Ce constat, issu d'une analyse épistémologique détaillée des structures inférentielles en question, est d'une importance capitale pour que les débats scientifiques relatifs à ces deux types de modèles soient menés de façon constructive et efficace.

Ainsi, cette thèse est d'abord et avant tout un projet de philosophie des sciences qui contribuera à structurer des réflexions épistémologiques plus larges par rapport à la classification, mais d'importantes contributions en philosophie de la biologie en ressortiront aussi. Afin de couvrir l'ensemble de cet espace théorique de manière efficace, cette thèse a été tissée à partir de trois parties partiellement indépendantes. Dans un premier temps (chapitres 1 et 2), j'explore la structure inférentielle de la phylogénétique par l'étude du courant de pensée qui a cristallisé cette façon de faire la systématique, soit la cladistique. Cela me permet de présenter concrètement le cohérentisme historique et de défendre son utilité dans le cadre d'une analyse épistémologique de pratiques classificatoires comme celles de la phylogénétique. Dans un second temps (chapitres 3 et 4), je vais à la rencontre de la tradition des espèces naturelles afin de démontrer que cette approche est inadéquate pour l'étude de pratiques complexes comme celles de la phylogénétique. Finalement, dans la troisième partie de la thèse (chapitres 5 à 7), j'analyse les pratiques phylogénétiques contemporaines dans le but d'améliorer la compréhension des enjeux présents au sein des débats impliquant l'opposition entre modèles réseaux et modèles arborescents. Cette dernière partie me permet aussi d'illustrer la puissance des analyses épistémologiques mobilisant le cohérentisme historique. Avant d'offrir un aperçu plus soutenu du contenu de chacun des chapitres qui composent le corps de cette thèse, quelques précisions sémantiques et méthodologiques s'imposent.

Tout d'abord, je dois expliquer mon utilisation des différents termes utilisés pour référer aux disciplines biologiques étudiées. La systématique est la discipline qui vise à construire des systèmes classificatoires biologiques, c'est-à-dire à décrire le monde biologique sur la base des

catégories d'appartenance des entités biologiques et de la mise en relation de ces catégories. L'ensemble des relations entre catégories engendre un *système*. Ces systèmes peuvent prendre la forme d'une hiérarchie de rangs dits taxonomiques (variétés, espèces, genre, familles, etc.), une approche largement majoritaire en biologie.

La phylogénétique est une manière spécifique de faire de la systématique. Il s'agit d'un ensemble de pratiques qui reconstruit des systèmes de relations entre catégories biologiques et qui interprète ces systèmes de façon historique. Les liens entre les catégories sont, dans cette optique, des liens généalogiques, de sorte que les systèmes phylogénétiques représentent les liens de parenté entre lignées biologiques. J'utilise l'expression « phylogénétique traditionnelle » pour référer à un sous-ensemble de pratiques phylogénétiques centrées sur la pensée arborescente décrite au chapitre 6. J'utilise le terme « cladistique » pour référer à la mouture de la phylogénétique traditionnelle s'étant développée sous l'impulsion de Hennig à partir des années 1960 et ayant fortement contribué à faire de la phylogénétique une discipline centrale à la biologie de l'évolution (Hennig 1966; Rieppel 2016).

J'utilise le terme « phénétique » pour référer à un autre ensemble de pratiques s'inscrivant au sein de la systématique. La phénétique est un courant de pensée s'étant développé simultanément (ou légèrement après) la cladistique et qui mettait de l'avant une systématique ancrée dans l'analyse quantitative des liens de similarité, ce qui contraste avec les analyses qualitatives alors dominantes. Au-delà de la quantification des données issues des analyses de similarité, la phénétique cherchait à construire des systèmes classificatoires *athéoriques* (Hull 1970; 2001). Un tel objectif irréalisable a bien sûr été rapidement rejeté, autant par les biologistes que par les philosophes, mais il convient de souligner que j'utilise tout de même l'expression « phénétique » pour désigner ce courant de pensée dans son ensemble. Cette précision est importante puisque mon utilisation du terme en question est différente de l'utilisation qu'en font certains et certaines biologistes. Le terme « phénétique » peut aussi être utilisé pour référer aux méthodes d'analyses quantitatives de la similarité qui témoignent de l'influence de la phénétique (le courant de pensée) sur l'ensemble de la systématique.

Je dois souligner une autre opposition entre le vocabulaire philosophique et celui des biologistes. Dans le contexte de cette thèse et en cohérence avec une perspective philosophique sur la question, je considère que la classification réfère à toute entreprise qui décrit les phénomènes en effectuant des découpages (qui mènent à l'identification de catégories) et en discriminant les entités (celle-ci fait partie de cette catégorie, celle-là, non), puis en reconstruisant des systèmes de relation incorporant les catégories ainsi identifiées (Tort 1989). La classification peut donc mener à la création de systèmes dits hiérarchiques, en ce sens que les catégories y sont emboîtées au sein de catégories plus inclusives, et ce, à répétition. Les classifications biologiques opèrent généralement de la sorte : les organismes forment des espèces; les espèces, des genres; les genres, des familles; les familles, des ordres; les ordres, des classes; etc. Une hiérarchie de rangs taxonomiques est ainsi formée. Cependant, la simple description des organismes sur la base de leur espèce d'appartenance est une opération classificatoire, au sens retenu dans cette thèse. Similairement, la mise en relation d'espèces, même si elle est faite *sans formation de catégories de rangs supérieurs*, est considérée ici comme une entreprise classificatoire puisqu'il s'agit de la création d'un système de relations (non hiérarchique). Au contraire, les biologistes ont l'habitude d'utiliser le terme « classification » pour référer exclusivement à la formation de systèmes hiérarchiques. L'analyse épistémologique rigoureuse requiert une acception plus inclusive de la notion de classification, sous peine de passer outre une partie significative des phénomènes d'intérêt.

Dans cette thèse, le terme « épistémologique » est utilisé pour référer aux discours et aux pratiques qui s'intéressent à la construction de connaissances. Le terme « épistémique » réfère quant à lui aux pratiques et aux discours de construction de connaissances. C'est donc dire que l'épistémologie est épistémique, puisqu'elle vise à construire des connaissances relatives aux entreprises de connaissance, mais tout projet épistémique n'est pas épistémologique. Le terme « métaphysique » réfère à tout discours relatif à ce que les choses *sont*. J'inclus dans ces discours ceux qui, comme le mien, considèrent que la seule métaphysique possible est celle qui s'intéresse à ce que les discours gnoséologiques nous disent au sujet de ce que les choses sont. En ce sens, je fais de la métaphysique des sciences lorsque je m'intéresse, par exemple, aux croyances véhiculées par les pratiques phylogénétiques à l'égard des entités biologiques et des relations

qu'elles entretiennent. Cette utilisation spécifique du terme « métaphysique » se rapproche de ce que les anthropologues culturels appellent une cosmologie, soit l'ensemble des entités (choses et processus) qu'une culture (incluant les cultures scientifiques) considère comme existant effectivement (Latour 2015). J'utilise le terme « ontologie » pour référer à l'inventaire des entités qui composent une métaphysique ou à l'étude du contenu associé à chacune de ces entités.

Quelques précisions méthodologiques s'imposent également. D'abord, il convient de souligner que le cohérentisme historique est une approche de l'étude des discours de connaissances qui invite à la pluridisciplinarité. La complexité des dynamiques scientifiques gagne ultimement à être approchée par l'entreprise de regards pluriels qui captent différentes projections de ce phénomène : les perspectives anthropologique, historique, sociologique, politologique, épistémologique sont autant d'angles d'approche qui informent sur la construction graduelle des connaissances et qui permettent la comparaison entre différents états de connaissance. Or, cette thèse est bien une étude de philosophie des sciences. En ce sens, je me limite à l'analyse qualitative du discours observable au sein des publications avec révision par les pairs qui composent le corpus de biologie de l'évolution étudié. Mon analyse qualitative est toutefois informée par des considérations historiques nécessaires à la bonne compréhension des liens entre les croyances qui animent le discours étudié. Ainsi, je mobilise des informations historiques à l'égard de la phylogénétique, et plus largement de la biologie de l'évolution, mais les contributions que cette thèse apporte au milieu de la recherche sont bien de nature philosophique plutôt qu'historique.

Toujours dans une perspective méthodologique, je dois préciser que la méthode argumentative préconisée, particulièrement dans le cadre des réflexions en philosophie des sciences (chapitres 1 à 4), se veut audacieuse. En effet, je prends fortement position par rapport à la tradition des espèces naturelles et du fondationnalisme. Cette manière d'approcher ces sujets vise dans un premier temps à explorer en profondeur le cohérentisme et l'éliminativisme à l'égard des espèces naturelles. Dans un second temps, j'espère que mes arguments intransigeants sauront susciter des réponses constructives de la part des tenants et des tenantes des postures philosophiques auxquelles mon propos s'oppose, de sorte à engendrer ou à continuer des dialogues édifiants.

Sur le plan linguistique, il convient de noter que cette thèse est rédigée de sorte à favoriser l'inclusivité. Pour ce faire, j'adopte une pluralité de stratégies, comme recommandé par l'Office québécois de la langue française¹. J'utilise notamment des formulations de genre neutre lorsque possible. Quand je dois féminiser des termes, j'utilise leur forme complète si les formes masculines et féminines ont des prononciations différentes. Si, au contraire, les deux formes se prononcent de la même façon (comme *impliqué* et *impliquée*), j'utilise le point médian (*impliqué-e*) pour rendre les deux genres visibles sans trop alourdir le texte. Je m'excuse auprès des personnes qui ne se sentent représentées par aucun de ces deux genres et j'espère que l'utilisation de termes neutres saura minimiser l'impact de cette situation liée à la dichotomisation des genres et des sexes humains inscrite dans la langue française (et je me tiendrai à l'affût des innovations linguistiques pertinentes, comme l'introduction du pronom *iel* dans le dictionnaire [*Le Robert*], que j'ai pu observer [très] tardivement dans le processus de rédaction du présent document).

Enfin, je présente la phylogénétique, tout au cours de cette thèse, comme étant une discipline classificatoire plutôt qu'historique. Ce choix peut surprendre les personnes familières avec ce champ d'études, la phylogénétique étant régulièrement présentée comme une discipline historique (Cleland 2002; Lecointre et Huneman 2020; O'Malley 2016; Velasco 2013). Au chapitre 7, je démontre que la dimension classificatoire de la discipline prime sur sa dimension historique. En bref, il s'agit de souligner que la phylogénétique érige des systèmes de relations entre unités taxonomiques et qu'une interprétation historique de ces systèmes peut suivre leur construction. Cette reconstruction historique peut cependant être omise (et l'est régulièrement), alors que la construction de systèmes d'affinités entre unités taxonomiques, basée sur des analyses de similarité, est le pain quotidien des phylogénéticiens et des phylogénéticiennes. En ce sens, la phylogénétique est d'abord une discipline classificatoire et ensuite une discipline partiellement historique.

¹ https://bdl.oqlf.gouv.qc.ca/bdl/gabarit_bdl.asp?id=3912

Toutes ces spécifications relatives à mon discours et à ma méthode ayant été explicitées, il convient de présenter le contenu des chapitres avant de se lancer à pieds joints dans l'analyse. Au premier chapitre de la thèse, je présente le problème de la coordination et son incarnation en phylogénétique. Le problème de la coordination est un phénomène épistémique qui a été théorisé au tournant du 20^e siècle. Ce phénomène, longtemps considéré comme se rapportant aux pratiques métrologiques, est le point de départ des réflexions associées au cohérentisme historique (Chang 2004; Tal 2016; van Fraassen 2008). La démonstration du fait que le problème de la coordination s'incarne au sein de la structure épistémique de la phylogénétique justifie l'utilisation du cohérentisme historique dans le cadre d'analyses épistémologiques de cette discipline puisque ce courant de pensée s'articule autour du problème en question.

Le problème de la coordination traduit de façon précise la charge théorique des observations. En effet, la coordination est ce processus par lequel on associe des éléments symboliques des sciences (quantités théoriques, hypothèses, modèles, etc.) avec des phénomènes concrets. Une telle coordination requiert des procédures d'observation : pour dire, par exemple, qu'un phénomène a une température de 100 °C, encore faut-il pouvoir observer la température (avec précision). Lorsqu'on réalise de telles observations, on coordonne un aspect du phénomène (le mouvement de ses particules constitutives) avec une quantité théorique (la température). Or, pour réaliser de telles observations avec un degré minimal de précision, il faut accepter de manière apriorique des croyances (comme celle d'un lien entre volume, quantité, pression et température qui permet de confectionner des thermomètres et d'avoir confiance en leur utilité) qui ne peuvent être vérifiées qu'une fois que les observations sont effectuées.

Ce problème s'incarne en phylogénétique. La structure inférentielle de cette discipline, dans son ensemble, vise à établir la teneur des liens phylogénétiques au sein de monde vivant. Or, pour reconstruire les systèmes de parenté entre entités biologiques, il faut accepter ce qui semble être au premier regard une circularité justificative : la phylogénétique, qui vise notamment à fournir un contenu empirique à la théorie de l'évolution par voie de sélection naturelle, doit accepter de façon apriorique le lien entre similarité et phylogénie, un lien qui est au cœur de ladite théorie (par l'entremise de la notion d'hérédité). Cependant, j'explore aussi, dans ce chapitre, l'histoire de la cladistique par l'entremise de la notion d'homologie, de sorte à

démontrer que cette circularité est illusoire. Les justifications empiriques, confrontées au problème de la coordination, tirent leur légitimité de pratiques scientifiques passées. Il s'ensuit que la structure épistémique de la discipline détient une dimension diachronique qui incite à utiliser un modèle hélicoïdal pour la représenter (plutôt qu'un cercle). En résumé, le premier chapitre me permet de présenter le problème de la coordination, la structure inférentielle de la phylogénétique, l'incarnation du premier au sein de la seconde et la dimension diachronique de l'ensemble.

Le second chapitre de la thèse explore les différentes approches philosophiques ayant traité du problème de la coordination, en lien avec les pratiques métrologiques et donc avec les descriptions quantitatives des phénomènes, afin d'utiliser ce corpus de textes pour informer une philosophie de la classification. Je présente d'abord le conventionnalisme (Mach 1896; Poincaré 1898; Reichenbach 1927), première tentative pour rendre compte philosophiquement des complications épistémiques liées au problème de la coordination. Cette approche considère les phénomènes de coordination dans une perspective anhistorique. Dans cette optique, la circularité associée à la coordination menace la validité des justifications empiriques : l'observation empirique qui est censée soutenir les théories scientifiques requiert l'acceptation apriorique d'éléments inscrits au cœur de ces théories. Afin de maintenir la validité de l'empirisme (la posture selon laquelle l'expérience serait la seule ou la principale source de légitimité des connaissances), le conventionnalisme, notamment sous la plume de Reichenbach, a suggéré que la circularité associée à la coordination serait un épiphénomène. La structure justificative des sciences (au sein de laquelle on observe ladite circularité) reposerait sur la correspondance entre le langage (ou la possibilité du langage) et les phénomènes. Ainsi, la circularité serait interne aux processus de construction de connaissances, mais l'ensemble de ces processus pourrait tout de même s'appuyer de manière absolue sur quelque chose qui lui est externe, soit la relation langage-phénomènes, de sorte à briser la circularité épistémiquement menaçante. En ce sens, l'empirie demeure *une fondation* pour la connaissance scientifique, un appui inébranlable accessible par le langage.

La deuxième moitié du 20^e siècle a été le théâtre d'une série de remises en question de la possibilité d'établir l'expérience comme une fondation absolue à la connaissance (Foucault 1966;

Quine 1961; Sellars 1956). Une posture opposée, à laquelle on réfère par l'étiquette « cohérentisme », a donc été développée, notamment en philosophie des sciences. L'une des moutures du cohérentisme ayant eu un impact significatif, du moins en philosophie de la mesure, est le constructivisme (Latour 2003; Tal 2012). Cette approche considère que les connaissances scientifiques sont construites par la communauté scientifique plutôt que découvertes à partir d'un monde qui serait indépendant de la cognition humaine. La validité des connaissances ne s'appuierait donc pas sur l'expérience, mais sur les normes sociales qui forgent le rapport à l'expérience et les différents aspects symboliques des pratiques scientifiques.

Le constructivisme a le défaut, selon ses critiques, de proscrire la possibilité de penser la normativité ou le progrès scientifique : la validité des connaissances étant issue des dynamiques propres à une communauté, toute connaissance soutenue par une communauté doit être considérée sur un pied d'égalité. Le constructivisme a cependant permis d'intégrer un constat extrêmement important au sein de la philosophie des sciences : les dynamiques de justification scientifique sont irréductibles à une simple correspondance entre phénomènes et théories puisque les théories sont sous-déterminées par les observations et déterminées par un ensemble complexe et hétérogène de facteurs. Les textes que j'associe au cohérentisme historique ont été écrits dans l'espoir explicite de réconcilier la possibilité de théoriser philosophiquement le progrès et la normativité scientifiques avec le constat selon lequel l'expérience est inapte à justifier seule l'acceptation de croyances, d'hypothèses, de modèles et de théories scientifiques. La deuxième partie du chapitre présente de façon détaillée trois approches que j'associe au cohérentisme historique, soit celles de Chang, de Tal et de van Fraassen, pour en extraire des outils d'analyse épistémologique applicables à l'étude des pratiques classificatoires au sein desquelles le problème de la coordination s'incarne. La conclusion du chapitre offre un résumé des apports du cohérentisme historique en répertoriant les concepts principaux que j'applique à l'analyse des pratiques phylogénétiques aux chapitres 5 à 7.

En bref, le chapitre 2 démontre que le cohérentisme historique est une approche adéquate pour l'analyse épistémologique de tout système épistémique caractérisé par la présence du problème de la coordination. Les chapitres 3 et 4 argumentent que la tradition des espèces naturelles, au contraire, est structurée d'une manière qui la rend incapable d'embrasser

pleinement la complexité de pratiques classificatoires comme la phylogénétique. Le chapitre 3 est une présentation du corpus de textes portant sur les espèces naturelles en philosophie des sciences, alors que le chapitre 4 est une formulation de mon argument éliminativiste selon lequel l'épistémologie des pratiques classificatoires devrait éviter d'être centrée sur la notion d'espèce naturelle.

Si l'objectif central du chapitre 3 est d'explorer la tradition des espèces naturelles, j'en profite pour mettre de l'avant, suivant Hacking (1991), l'idée selon laquelle cette tradition est unitaire, du moins en philosophie des sciences. Plus précisément, je démontre que les principales théories des espèces naturelles ayant été articulées depuis la cristallisation de la tradition au 19^e siècle (Mill 1862; Whewell 1840) sont structurées sur le même modèle que je mets à jour et que j'appelle l'arche des espèces naturelles. En résumé, l'arche traduit un sentiment naturaliste véhiculé par la tradition des espèces naturelles (Ereshefsky et Reydon 2015) selon lequel les catégories scientifiques doivent rendre compte de divisions existant dans la structure du monde et réussissent à le faire (l'expression populaire, en anglais, pour traduire la chose est la suivante : *they carve the world at its joints*).

Cette posture naturaliste peut être décomposée en quatre éléments qui forment l'arche des espèces naturelles. Cette arche détient deux piliers. Un pilier représente la croyance selon laquelle certaines catégories scientifiques ont un statut épistémique particulier, notable et qui témoigne d'une hiérarchie parmi les catégories mobilisées par les êtres humains. L'autre pilier représente la croyance selon laquelle certaines catégories ont un statut métaphysique particulier (en raison de leur essence constitutive ou d'une structure causale particulière, par exemple). Les deux piliers sont solidifiés par une tige de métal ancrée aux piliers par des essences. Cette tige et ses ancrages représentent le lien logique entre les deux croyances tout juste présentées, un lien qui peut prendre deux formes : soit le fait métaphysique au sujet des catégories est utilisé pour expliquer le fait épistémique, soit le fait épistémique est utilisé pour inférer le fait métaphysique (et parfois, les deux formes sont présentes au sein d'une théorie des espèces naturelles). Ces trois éléments (les deux piliers et l'essence) soutiennent l'arc de l'arche, soit la croyance selon laquelle l'épistémologie doit se doter d'une ou de plusieurs théories des espèces naturelles afin d'analyser les pratiques classificatoires en sciences. Les différentes analyses réalisées au chapitre 3

démontrent que ce modèle s'incarne dans une grande diversité de théories des espèces naturelles qui représentent les principales familles d'approches présentes au sein de la tradition, soit le pluralisme (Boyd 1999a; Dupré 1993), le monisme essentialiste (Bird 2009; Devitt 2008; LaPorte 2004) et le monisme non essentialiste (Khalidi 2013; Slater 2015).

Au chapitre 4, j'articule un argument éliminativiste selon lequel l'épistémologie des pratiques classificatoires gagne à faire l'économie des théories des espèces naturelles en remettant en question à la fois l'arc et ses deux piliers. Mon argument est d'abord formulé sous la forme d'une réponse à un article de Brigandt (2009) au sein duquel il défend explicitement l'utilité épistémologique de la notion d'espèce naturelle. Son argument comporte deux éléments : premièrement, le fait métaphysique impliqué dans les phénomènes associés à des espèces naturelles explique l'utilité de l'identification d'espèces naturelles; deuxièmement, il est nécessaire d'avoir des concepts associés à des espèces naturelles pour incarner la connaissance de façon efficace. En bref, mes réponses consistent à souligner qu'au sein de ces deux arguments, la notion d'espèce naturelle est superflue : d'abord, l'identification d'espèces naturelles est réalisée après l'analyse épistémologique de pratiques scientifiques, de sorte qu'elle est accessoire à ces analyses; ensuite, s'il est fort utile d'avoir des concepts pour simplifier la communication et incarner nos connaissances par rapport à certains objets d'étude, ce constat est valide même pour des concepts indépendants de toute espèce naturelle (comme en témoigne l'utilisation de concepts explicitement considérés par les scientifiques comme des fictions utiles).

À ces deux réponses au texte de Brigandt (2009), j'ajoute deux arguments supplémentaires pour ébranler l'architecture de la tradition des espèces naturelles. Premièrement, je démontre que les théories des espèces naturelles sont des mécanismes de légitimation des pratiques scientifiques plutôt que des cadres d'analyse épistémologiques pour appréhender les pratiques classificatoires. Deuxièmement, en sciences, les pratiques classificatoires traduisent des toiles épistémiques complexes au sein desquelles les catégories scientifiques ne sont qu'un élément d'intérêt parmi d'autres. En focalisant le propos sur ces catégories, les théories des espèces naturelles contournent donc une partie significative des pratiques qui devraient pourtant être l'objet d'étude d'une épistémologie de la classification.

En plus de formuler ces arguments qui justifient l'adoption d'une posture éliminativiste à l'égard des espèces naturelles, le chapitre 4 présente deux autres postures éliminativistes, soit celles de Ludwig (2018), fortement inspirée de Hacking (2007a), et de Brigandt (2020). Le tout forme un ensemble d'arguments qui renverse le fardeau de la preuve sur les tenants et tenantes des théories des espèces naturelles : il leur revient dorénavant de démontrer l'utilité épistémologique des théories des espèces naturelles. Ce dernier point ajoute une précision importante à ma posture éliminativiste : l'objectif est bien d'engendrer un dialogue avec les théoriciens et théoriciennes des espèces naturelles afin de conserver exclusivement les (nombreux) éléments théoriques utiles qui ont été développés au sein de la tradition des espèces naturelles, tout en considérant sérieusement la possibilité de structurer l'épistémologie des pratiques classificatoires autrement qu'en focalisant les analyses sur la notion d'espèce naturelle.

Ainsi, les chapitres 3 et 4 me permettent d'identifier un modèle, soit l'arche des espèces naturelles, pour l'analyse des pratiques classificatoires en sciences et d'en démontrer les faiblesses : 1) la notion d'espèce naturelle est moins importante pour l'analyse épistémologique des pratiques classificatoires que l'arche des espèces naturelles le considère (et *contra* Brigandt 2009); 2) le modèle est plus utile pour légitimer les pratiques classificatoires que pour les analyser; 3) l'accent mis sur les catégories scientifiques est insuffisant pour une analyse épistémologique des pratiques classificatoires. Les chapitres 5, 6 et 7, en plus de mener à d'importants constats en philosophie de la biologie, illustrent le fait que le cohérentisme historique est apte à analyser les pratiques classificatoires en comblant les lacunes des théories des espèces naturelles.

Le chapitre 5 présente la transition aux données moléculaires qui caractérise la biologie de la deuxième moitié du 20^e siècle et qui a largement contribué à définir la phylogénétique d'aujourd'hui. En ce sens, l'exploration de cette transition et de son impact sur la structure inférentielle de la discipline est d'une importance capitale pour la compréhension des pratiques phylogénétiques contemporaines. Je mobilise donc les outils du cohérentisme historique pour analyser les deux itérations de la phylogénétique, celle centrée sur les données morphologiques et celle qui mobilise des données moléculaires. Afin de pouvoir comparer une telle analyse avec le travail qui pourrait être effectué en mobilisant l'arche des espèces naturelles, l'analyse est limitée à la notion d'homologie.

Le constat principal de ce chapitre est que la structure inférentielle de la phylogénétique est restée largement intacte malgré l'importante transition sur le plan des données, comme cela est souvent souligné (Inkpen et Doolittle 2016; O'Malley 2016; Sapp 2009). Cependant, une analyse épistémologique de la notion d'homologie qui va au-delà des définitions et des aspects théoriques de la chose permet de mettre à jour une différence importante entre l'homologie moléculaire et l'homologie morphologique. En effet, les modèles qui sous-tendent l'analyse de la similarité requise pour postuler et confirmer des relations d'homologie divergent significativement. Les modèles d'analyse de la similarité morphologique sont indépendants de toute hypothèse relative à la transformation des traits, alors que de telles hypothèses relatives aux transformations subies par les séquences génétiques sont nécessaires pour la pondération de la similarité moléculaire. Le tout témoigne de la validité du modèle hélicoïdal que j'associe au cohérentisme historique : l'itération moléculaire de la phylogénétique, bien que très similaire à l'itération précédente sur laquelle elle s'appuie, se différencie sur certains aspects théoriques relatifs à l'opérationnalisation de la discipline; le cercle justificatif propre à chaque itération analysée indépendante de l'autre s'étire donc sur un axe temporel qui les relie, ce qui engendre une hélice, et dans ce cas-ci, l'introduction de considérations évolutives au sein des modèles d'analyse de la similarité.

Le chapitre 6 amorce l'analyse comparative des pratiques phylogénétiques centrées sur les modèles réseaux et les modèles arborescents, ces modèles qui caractérisent la phylogénétique contemporaine. Cette tâche requiert notamment la présentation d'un autre développement historique relatif à l'utilisation des données moléculaires. Il s'agit de la montée en puissance de la phylogénomique, soit l'adaptation des pratiques phylogénétiques au traitement de données massives issues, entre autres, de l'amélioration des techniques de séquençage d'ADN s'étant accélérée au tournant du 21^e siècle et de l'augmentation de la puissance calculatoire disponible pour les biologistes. J'y présente donc la phylogénomique, mais je centre surtout mon propos sur la pensée arborescente et son incarnation au sein des pratiques contemporaines (notamment au sein des programmes de recherche phylogénomiques).

La pensée arborescente réfère au cadre théorique dominant la phylogénétique de la deuxième moitié du 20^e siècle. Il s'agit d'un cadre théorique centré sur les modèles arborescents

et leur utilisation pour l'inférence des liens de parenté entre unités taxonomiques. J'analyse cette toile épistémique à trois niveaux distincts : premièrement, j'analyse une formulation explicite de la pensée arborescente par O'Hara (1997); deuxièmement, j'explore l'incarnation de cette pensée au sein d'un texte programmatique, soit un texte de Delsuc, Brinkmann et Philippe (2005) qui cherche à décrire et structurer la phylogénomique naissante en adhérant à la pensée arborescente; troisièmement, je me penche sur un débat phylogénétique concret portant sur le placement phylogénétique de *Methanonatronarchaeia*, un phylum d'archées récemment découvert (Aouad et al. 2018, 2019; Sorokin et al. 2017, 2018, 2019). Ce débat témoigne du fait que les éléments importants de la pensée arborescente tels qu'on les observe au sein de textes du début du siècle (Delsuc, Brinkmann et Philippe 2005; O'Hara 1997) sont toujours actualisés au sein des pratiques les plus récentes.

Cette analyse à plusieurs niveaux d'abstraction (allant de la description épistémologique de la pensée arborescente jusqu'à son incarnation au sein des pratiques concrètes) des pratiques phylogénétiques que je décris comme étant traditionnelles (en raison de leur association à la pensée arborescente) me permet de formuler trois constats. Ces constats contribuent à la bonne compréhension du phénomène qu'est la phylogénétique traditionnelle, mais ils mettent aussi la table pour l'analyse comparative entre la phylogénétique traditionnelle et celle centrée sur les réseaux.

Le premier constat concerne une série d'amalgames véhiculés par la pensée arborescente qui teinte les pratiques phylogénétiques. J'utilise le terme *amalgame* pour référer à des associations d'idées présentées comme étant nécessairement liées alors qu'une analyse logique historiquement informée témoigne de l'indépendance de ces idées. De tels amalgames sont nécessaires pour simplifier les pratiques scientifiques (qui ne peuvent pas constamment remettre en question toutes les associations d'idées mobilisées). Cependant, ils peuvent être nuisibles d'au moins deux façons : ils nuisent à la bonne description épistémologique des pratiques scientifiques en présentant des idées indépendantes comme si elles étaient nécessairement liées; parce qu'ils favorisent de mauvaises descriptions épistémologiques, les amalgames nuisent aussi à la réalisation de débats scientifiques sains. Comme je le démontre aux chapitres 6 et 7, il est en effet important d'éviter de confondre une remise en question du modèle arborescent avec une remise

en question de l'ensemble du bagage théorique évolutionniste qui est associé à la pensée arborescente.

Le premier amalgame consiste à associer le modèle arborescent au rejet de la pensée du groupe et de la pensée développementale. En effet, O'Hara présente la pensée arborescente, qui utilise le modèle arborescent à la fois pour réaliser des inférences phylogénétiques et représenter l'histoire évolutive du monde vivant, en opposition à ces deux pensées. La pensée du groupe considère les unités taxonomiques comme des occurrences indépendantes de leur rang taxonomique (les espèces seraient des occurrences indépendantes du rang *espèce*, par exemple), alors que la pensée arborescente implique une pensée évolutive selon laquelle les espèces seraient liées généalogiquement les unes aux autres. La pensée du développement assimile progrès (ou complexification) avec évolution, alors que la pensée arborescente rejette cette association d'idées. S'il est vrai que les pratiques phylogénétiques contemporaines du texte d'O'Hara utilisent le modèle arborescent et rejettent les pensées du groupe et du développement, je souligne que le modèle arborescent est tout de même (historiquement et logiquement) compatible avec ces deux pensées. Ainsi, l'association entre le modèle arborescent et le rejet de ces deux pensées est un amalgame d'idées indépendantes.

Le deuxième amalgame consiste à lier la pratique de la phylogénétique au modèle arborescent. Comme en témoigne la phylogénétique centrée sur les réseaux, cette association d'idées gagne à être décomposée. De plus, l'histoire de la systématique nous apprend que le modèle arborescent peut être utilisé sans que les systèmes d'affinités entre unités taxonomiques soient interprétés de manière phylogénétique. L'indépendance entre modèle arborescent et phylogénétique est donc mutuelle.

Le troisième amalgame traduit le lien entre l'étude de l'évolution et l'étude des lignées d'organismes ou de gènes. En effet, la phylogénétique, en tant qu'étude des dynamiques de formation de lignées, est considérée comme un champ disciplinaire central de la biologie de l'évolution (Kluge 2002; Wiley et Lieberman 2011). Ce qui sous-tend cette idée est que les dynamiques évolutives impliquent nécessairement des lignées, c'est-à-dire des successions linéaires ou quasi linéaires d'entités biologiques comme des organismes ou des gènes. Or, le fait

que de telles lignées soient présentes n'implique *pas nécessairement* que les phénomènes évolutifs gagnent à être compris par l'entremise de l'étude de ces lignées, comme en témoigne une série de travaux récents (Charbonneau 2014; Papale 2021; Roughgarden 2020). Ainsi, l'association d'idées selon laquelle l'étude de l'évolution requiert l'étude des dynamiques de formation de lignées est bel et bien un amalgame véhiculé tout autant par la phylogénétique centrée sur les réseaux (comme démontré au chapitre 7) que par la phylogénétique traditionnelle.

Ces trois amalgames composent le premier constat, soit celui selon lequel la pensée arborescente est un complexe de croyances qui gagne à être décomposé et considéré dans une perspective historique afin de favoriser l'analyse épistémologique et la bonne tenue des débats scientifiques. Le deuxième constat est celui de la présence d'un principe ontologique entretenu par la phylogénétique traditionnelle. Cette notion, inscrite au sein du cohérentisme historique (Chang 2004), réfère à des croyances non justifiées (et souvent non justifiables) qui structurent un champ disciplinaire de manière importante. La phylogénétique est traversée par ce que j'appelle le principe d'histoire évolutive unique, soit la croyance selon laquelle l'histoire du monde biologique pourrait être représentée à l'aide d'un seul schème représentatif qui pourrait résumer l'ensemble des événements significatifs (sans faire appel à la notion de principe ontologique, Baptiste 2007 décrit ce phénomène culturel en détail). Au sein de la phylogénétique traditionnelle, ce principe ontologique forge le rapport des praticiens et praticiennes au modèle arborescent : celui-ci est soutenu par une vision *réaliste* et *moniste* qui mène à croire que l'histoire du monde vivant peut être représentée de façon complète par une seule représentation arborescente, soit l'Arbre du vivant.

Le troisième constat découle du second. En effet, la présence de ce principe ontologique et de la posture réaliste et moniste qui l'accompagne témoigne d'un rapport aux connaissances bien précis. Les phylogénéticiens et phylogénéticiennes adoptent, implicitement (Sorokin et al. 2019, par exemple) ou explicitement (Kluge 1998, par exemple), une posture épistémologique fondationnaliste. Ils et elles considèrent que les modèles et théories peuvent se conformer à la réalité telle qu'elle existe indépendamment de nos pratiques épistémiques, et donc que l'évaluation de cette correspondance serait le moteur principal du progrès scientifique. Cette

posture fondationnaliste est particulièrement bien exemplifiée au sein de la philosophie de la preuve intégrale (Kluge 1998; Rieppel 2009) qui cherche explicitement à faire l'économie des contraintes théoriques. Cette logique est aussi exemplifiée par le débat concernant *Methanonatronarchaeia* que j'explore dans le chapitre 6. Alors même que les auteurs et autrices débattent d'aspects méthodologiques de la pratique qui portent sur le choix des données et la façon de les analyser, ils et elles expriment leur espoir que l'accumulation de nouvelles données viendra clore le débat (Aouad et al. 2019; Sorokin et al. 2019).

Il va sans dire que cette tendance à adopter le fondationnalisme est incompatible avec les conséquences épistémiques du problème de la coordination. L'observation est nécessairement chargée théoriquement et les débats méthodologiques portant sur les données sont influencés par la collecte de nouvelles données, mais ne peuvent être tranchés que par des décisions prises à l'égard de procédures qui sous-tendent l'observation. Ainsi, afin de favoriser le progrès scientifique ancré dans une bonne compréhension des dynamiques des débats scientifiques, je suggère que la phylogénétique gagnerait à adopter un regard sur ses propres processus de construction de connaissances qui correspond plus adéquatement aux postures philosophiques contemporaines. Les outils du cohérentisme historique se prêtent merveilleusement à la tâche, comme en témoignent les analyses réalisées aux chapitres 6 et 7 de cette thèse.

Le chapitre 7 propose une analyse comparative des phylogénétiques centrées sur les arbres et les réseaux. Pour ce faire, il faut nécessairement réaliser une analyse détaillée de la phylogénétique centrée sur les réseaux, ce que j'effectue dans la première partie du chapitre. En phylogénétique, les modèles réseaux sont principalement mobilisés afin de représenter les phénomènes évolutifs qui mènent à la réticulation des lignées, soit la présence de liens transversaux entre lignées que les modèles arborescents représenteraient comme étant strictement divergentes. En effet, là où les modèles arborescents représentent l'histoire évolutive des unités taxonomiques en priorisant la bifurcation et la divergence (une lignée laisse place à deux lignées descendantes qui divergent toujours plus l'une de l'autre), les modèles réseaux permettent à deux lignées de conjuguer leurs efforts pour en engendrer une troisième, leur descendante. Une panoplie de phénomènes biologiques justifient l'utilisation de modèles réseaux (les transferts génétiques entre lignées, l'hybridation des lignées, l'endosymbiose, etc.). Ainsi, des

différences importantes existent entre la phylogénétique centrée sur les réseaux et celle centrée sur les arbres : elles représentent l'histoire évolutive différemment (avec ou sans réticulation) et elles accordent une importance différente à la divergence, un principe théorique au cœur de certaines interprétations de la théorie de l'évolution par voie de sélection naturelle (dont celle de Darwin).

Cependant, certaines distinctions sont plus nuancées. Si la phylogénétique centrée sur les réseaux remet en question la dominance du modèle arborescent, elle mobilise tout de même des modèles arborescents au sein des inférences qui mènent à la reconstruction de réseaux. De plus, elle ouvre la porte à la réticulation, mais considère toutefois que celle-ci est l'exception et non la règle. En ce sens, si deux hypothèses sont également soutenues par un jeu de données, celle présentant le moins de réticulation devrait être priorisée. Par ailleurs, la posture réaliste soutenue par le principe d'histoire évolutive unique apparaît aussi au sein de la phylogénétique centrée sur les réseaux, bien qu'elle tende à y être moins importante.

Finalement, une série de liens de similarité très forts unissent les deux itérations phylogénétiques : les analyses de similarité moléculaire et morphologique qu'elles mobilisent sont les mêmes et sont utilisées de la même façon pour inférer les liens phylogénétiques; les deux s'articulent autour de l'importance des lignées; les deux itérations de la phylogénétique sont des entreprises éminemment classificatoires. L'ensemble de cette analyse comparative est représenté de manière synoptique dans le tableau 7.1.

Cette analyse comparative est, en soi, une contribution importante à une meilleure compréhension des pratiques phylogénétiques contemporaines. Elle permet aussi d'informer différentes réflexions épistémologiques directement liées à ces deux itérations de la phylogénétique. Ainsi, la deuxième partie du chapitre 7 présente cinq débats philosophiques qui ont des conséquences importantes pour les pratiques scientifiques concernées. Les trois premiers débats sont présents au sein du corpus de textes étudié. Les deux autres sont des débats qui brillent par leur absence au sein de ce corpus. Je considère toutefois que ceux-ci méritent de l'attention des philosophes et biologistes intéressé·e·s par la phylogénétique ou, plus largement, par la biologie de l'évolution.

Le premier débat concerne la compatibilité entre les modèles réseaux et la théorie de l'évolution par voie de sélection naturelle. Les réseaux ont en effet été considérés comme entrant en conflit avec les principes centraux du cadre conceptuel dominant la biologie de l'évolution (DeSalle et Riley 2020; Forterre 2012; Raoult et Koonin 2012). Je démontre, en mobilisant des travaux réalisés ailleurs (Baptiste et Papale 2021), que la posture selon laquelle les modèles réseaux, bien qu'ils remettent en question certains aspects de la conception traditionnelle de la théorie (néo-)darwinienne de l'évolution, s'inscrivent en continuité avec ce qui les précède. Je profite de l'occasion pour réitérer l'importance des analyses épistémologiques qui décomposent avec précision les schèmes de pensée afin de pouvoir prendre position adéquatement par rapport aux inévitables transformations de ces schèmes de pensée.

Le second débat porte sur le réalisme en phylogénétique. À la suite de Velasco (2012), je démontre d'abord que des personnes adoptant une posture réaliste par rapport aux connaissances phylogénétiques devraient être réalistes à l'égard des réseaux phylogénétiques plutôt qu'à l'égard de l'Arbre du vivant. Ensuite et surtout, je souligne que la posture réaliste perd de son importance face aux constats épistémologiques liés au problème de la coordination et à son incarnation en phylogénétique.

Le troisième débat porte sur l'utilité épistémique des modèles réseaux et arborescents. Mon constat à ce sujet est celui d'une complémentarité de ces deux familles de modèles. En phylogénétique, leur utilité respective empêche, du moins à court terme, de rejeter l'un des deux modèles de façon absolue. En effet, les modèles réseaux offrent des représentations plus précises des phénomènes d'évolution, mais leur utilisation requiert une puissance calculatoire (beaucoup) plus grande qui justifie, dans bien des cas, l'utilisation de modèles arborescents (Morrison 2014a; Wen et al. 2018). De surcroît, l'utilisation de modèles réseaux implique souvent l'utilisation de modèles arborescents (comme démontré plus tôt dans le chapitre). Finalement, les représentations simplifiées associées aux modèles arborescents sont, dans bien des cas, suffisantes à l'étude des phénomènes d'intérêt (DeSalle et Riley 2020; Velasco 2012).

Les quatrième et cinquième débats concernent des éléments de similarité entre les deux itérations phylogénétiques. Le quatrième débat souligne les conséquences du fait que les deux

itérations reposent sur le troisième amalgame identifié au chapitre 6, soit le lien entre l'étude des lignées et l'étude de l'évolution. J'utilise un exemple, soit celui du gène chimérique *bla_{NDM}* et de la résistance à certaines substances antimicrobiennes (Haraoui 2021; Toleman et al. 2012), pour illustrer les limitations d'une biologie de l'évolution centrée sur l'étude des lignées.

Le cinquième débat concerne l'impératif classificatoire qui domine les descriptions du monde biologique (Lecointre et Huneman 2020), et tout particulièrement les descriptions mobilisant la nomenclature linnéenne établie sur des bases phylogénétiques. Plus précisément, je souligne que les entités biologiques sont généralement décrites sur la base de leur catégorie d'appartenance, et donc sur la base des points partagés entre les différents membres d'une catégorie. Or, un tel mode descriptif relègue à l'arrière-plan la description des particularités des individus, des différences interindividuelles. Ce constat implique nécessairement une perte de précision descriptive qui est particulièrement néfaste dans un contexte d'évolution : le cadre théorique darwinien a été tissé autour des différences interindividuelles qui sont le carburant de la sélection naturelle. Ainsi, la primauté de modalités descriptives qui nient ces différences interindividuelles au sein d'une biologie qui accorde une importance capitale aux dynamiques évolutives engendre une dissonance au sein de l'écosystème épistémique concerné. Je suggère donc d'accorder une plus grande importance aux descriptions quantitatives du monde vivant, sur la base de méthodes déjà disponibles au sein des pratiques phylogénétiques (que j'explore par l'entremise du texte de Parks et al. 2018). La résolution d'un tel débat, qui reste entièrement à définir, aura des conséquences importantes sur le développement de la phylogénétique et de la place de l'impératif classificatoire au sein de la biologie de l'évolution. J'ai aussi bon espoir que les réflexions entourant la question pourront éventuellement informer la place de l'impératif classificatoire bien au-delà de la communauté scientifique.

En résumé, cette thèse se veut une prise de position quant à la manière d'appréhender les pratiques classificatoires dans une perspective épistémologique. Je défends la validité des outils issus du cohérentisme historique et démontre leur puissance par l'analyse de pratiques phylogénétiques contemporaines. Le tout me permet de contribuer à la bonne compréhension de ces pratiques phylogénétiques et de faciliter les débats internes à la discipline concernée.

Cette thèse contribue notamment à une meilleure compréhension des bouleversements que la biologie de l'évolution connaît actuellement.

Chapitre 1 – Le problème de la coordination et la structure épistémique de la phylogénétique

All that you touch you Change.
All that you Change Changes you.
The only lasting truth is Change.
- Octavia Butler, dans *Parable of the Sower*

1.1 Introduction

L'objectif de ce chapitre est de démontrer que le problème de la coordination, un phénomène épistémique qui a structuré des réflexions en philosophie de la mesure tout au cours du 20^e siècle, est présent au cœur des pratiques classificatoires et, plus spécifiquement, en phylogénétique. Ce constat est d'une importance capitale pour toute analyse épistémologique de cette discipline. En effet, le problème de la coordination, lorsqu'il s'incarne, imprègne l'ensemble de la structure épistémique du champ gnoséologique auquel il est attaché, mettant en exergue la charge théorique de l'expérience. De fait, je cherche, dans ce qui suit, non seulement à mettre en lumière sa présence au sein de la phylogénétique, mais aussi à signaler sa portée et son importance.

Pour ce faire, je présente d'abord le problème de la coordination tel qu'il est appréhendé en philosophie de la mesure. Ensuite, je réarticule cette situation épistémique de sorte à la généraliser et à montrer qu'elle s'applique à la phylogénétique. Plutôt que de plonger immédiatement dans l'analyse de la phylogénétique contemporaine, je me penche par la suite sur des exemples historiques. Cela me permet, en plus d'illustrer le problème de la coordination, d'insister sur certains éléments théoriques et historiques de la phylogénétique qui permettent une meilleure appréciation des pratiques contemporaines. Finalement, ce chapitre permet de souligner le fait que la présence du problème de la coordination impose, d'un point de vue épistémologique, l'adoption d'une perspective historique pour analyser les structures épistémiques des pratiques scientifiques. Ce constat est fortement inspiré des travaux de Chang, Tal et van Fraassen que j'analyse en détail au chapitre 2.

1.2 Le problème de la coordination

Le problème de la coordination est étroitement lié à la justification empirique des théories scientifiques (Chang 1995, 2001, 2004; Tal 2012, 2013, 2015, 2016; van Fraassen 2008, 2012). Pour qu'une pratique métrologique soit considérée comme pouvant mesurer une quantité physique théoriquement délimitée, il faut pouvoir justifier, considérant une procédure P et une quantité X , la croyance selon laquelle la procédure P permet de mesurer la quantité X . La coordination est ce processus qui permet de relier un phénomène à une quantité théorique (ou vice versa) et qui répond à la question suivante : « What counts as a measurement of (physical quantity) X ? » (van Fraassen 2008, 116). Or, cette question, au premier regard, invite des réponses circulaires qui ébranlent toute tentative de fournir des bases absolues à la connaissance empirique. C'est en ce sens qu'il y a *problème* de la coordination.

Chang (2001, 2004, 2007) fournit une description limpide, en quatre étapes, de ce qu'il appelle le problème de la mesure associée à des lois (*the problem of nomic measurement*), synonyme du problème de la coordination² (Tal 2015) : 1- l'objectif est de mesurer une quantité X ; 2- la quantité X ne peut être observée directement par la perception humaine non assistée, ce qui requiert de l'inférer à partir d'une autre quantité, Y ; 3- afin d'effectuer cette inférence, il est nécessaire de se doter d'une loi qui exprime la fonction reliant X et Y telle que $X = f(Y)$; 4- pour tester empiriquement cette fonction, il faudrait pouvoir établir les valeurs de X et Y indépendamment de ladite fonction, mais la valeur de X est justement ce que l'on cherche à établir (Chang 2001). C'est donc dire que le fait même de mesurer une quantité implique une pétition de principe : il faut s'engager envers une relation déterminée afin de mesurer une quantité physique, alors que seul l'établissement de cette valeur nous permet de corroborer la relation en question.

Cette pétition de principe peut être illustrée par un exemple centré sur la notion de température et son étude au tournant du 19^e siècle. La thermométrie, à l'époque, était loin d'être aussi stable et standardisée qu'elle ne l'est aujourd'hui : « Quelle substance thermométrique

² L'expression préférée dans le cadre de cet ouvrage est bien « problème de la coordination » puisque celle-ci désigne avec plus de précision la forme générale du problème, qui s'applique notamment aux pratiques classificatoires.

devrait être utilisée? », « Quelle sorte de verre utiliser pour enfermer la substance thermométrique? », « Doit-on préconiser des thermomètres à pression fixe ou à volume fixe? », « Quelle échelle devrait être utilisée? », etc. sont tout autant de questions qui étaient alors d'actualité (Chang 2004). Dans un tel contexte d'instabilité métrologique, comment la validité empirique d'énoncés théoriques peut-elle être établie?

Prenons l'exemple de la loi de Gay-Lussac, selon laquelle la pression (P) et la température (T) varient inversement à volume et quantité fixes, tel que $P_1 \times T_1 = P_2 \times T_2 = k$, une constante³. Comment vérifier la validité *empirique* de cette fonction qui nous permettrait notamment de mesurer la température? Pour ce faire, il faut, entre autres choses, être muni d'un thermomètre, c'est-à-dire d'un outil permettant d'attribuer une valeur de température à un phénomène (van Fraassen 2008). Tout résultat empirique confirmant ladite loi est donc sujet aux mêmes critiques que celles visant les instruments mobilisés au sein des débats en thermométrie (Chang 2004). De surcroît, la confection de thermomètres au tournant du 19^e siècle, qu'ils soient fabriqués à base d'air, d'alcool ou de mercure, se faisait inmanquablement sur la base d'une relation entre volume, pression et température, cette même relation qui est au cœur de la loi de Gay-Lussac que les thermomètres permettent de vérifier (en admettant que des procédures pour la quantification de la pression et du volume soient aussi disponibles). La boucle est bouclée, de sorte que la vérification empirique d'énoncés impliquant la température semble en effet souffrir d'une circularité justificative inéluctable.

Cette situation illustre puissamment les difficultés épistémiques liées à la coordination de termes théoriques, comme celui de température, et de phénomènes concrets. Comme le souligne Chang, le problème de la coordination nous force à reconnaître que les observations, telles qu'elles sont réalisées par l'entremise de diverses procédures, impliquent nécessairement une charge théorique : « This circularity is probably the most crippling form of the theory-ladenness of observation. » (Chang 2001, 252). Il s'agit d'une situation handicapante (*crippling*) en ce sens que la circularité justificative est considérée comme étant nuisible sur le plan épistémique. En

³ La formulation complète de la loi des gaz parfaits, soit $PV = nRT$ n'a été établie que plus tard au 19^e siècle, en croisant les lois de Boyle-Mariotte et celle décrite dans le corps du texte.

effet, du point de vue de l'empirisme fondationnaliste, l'expérience devrait fournir les fondements absolus (qui sont eux-mêmes valables même en l'absence de justifications) pour la connaissance, les sciences y puisant leur légitimité; or, si toute expérience ne peut être reliée à la connaissance que par une structure circulaire, elle est de fait incapable de servir de fondement absolu à la connaissance.

Les auteurs contemporains et autrices contemporaines de la philosophie de la mesure insistent sur un fait important : cette circularité est illusoire (Chang 2004, 2012; Morrison 2009; Morrison et Morgan 1999; Tal 2016; van Fraassen 2008). Elle émerge non pas de la structure concrète des pratiques scientifiques, mais plutôt du traitement anhistorique de ces dernières. Lorsqu'un argument scientifique est analysé hors de son contexte sociohistorique, il témoigne sans aucun doute d'une circularité alarmante. En revanche, lorsqu'on le replonge dans son contexte, cette circularité révèle sa dimension diachronique, de sorte qu'un modèle hélicoïdal semble mieux adapté à la situation (Chang 2016) que le modèle circulaire préconisé par des auteurs tels que Reichenbach (1927). Un argument scientifique basé sur des données empiriques, pris indépendamment du reste de la discipline, est circulaire, mais sa puissance et sa valeur épistémiques sont en fait issues de son inscription au sein d'un processus adéquatement décrit par le modèle hélicoïdal au sujet duquel j'offre plus de détails dans le chapitre 2.

Cela étant dit, quel que soit le modèle adopté pour décrire les dynamiques épistémiques des pratiques visant la corroboration empirique de théories, le problème lié à la coordination est un obstacle de taille pour toute tentative de formuler une épistémologie empiriste et *fondationnaliste*, c'est-à-dire une épistémologie considérant l'expérience comme fondement *absolu* de la connaissance. Dans le prochain chapitre, j'explore les réponses que des philosophes de la mesure ont données à ce problème, notamment afin d'extraire des outils utiles pour l'analyse de la phylogénétique contemporaine et de détailler le modèle hélicoïdal mentionné précédemment. Cependant, il convient avant tout de démontrer que le problème de la coordination est bel et bien présent au sein de cette discipline, ce que je fais dans les trois prochaines parties (1.3, 1.4 et 1.5) en reformulant le problème et en mobilisant un exemple précis, soit la cladistique hennigienne.

1.3 Généralisation du problème de la coordination

Le problème qui nous intéresse a traditionnellement été considéré comme relevant de la philosophie de la mesure. Par exemple, Van Fraassen insiste sur le fait que la coordination relève de la mesure : « The theory would remain a piece of pure mathematics, and not an empirical theory at all, if its terms were not linked to *measurement* procedures. » (van Fraassen 2008, 115, italiques ajoutés). Chang, quant à lui, réfère à ce problème comme étant celui de la *mesure* associée à des lois (*problem of nomological measurement*). Cela dit, quelques années après avoir théorisé cette situation épistémique dans le contexte métrologique, Chang a aussi identifié la présence de dynamiques de justification similaires dans le cadre de pratiques classificatoires en chimie (Chang 2012, 2015). Dans ce qui suit, j'emboîte le pas en démontrant explicitement que le problème de la coordination s'incarne en phylogénétique et qu'il peut donc servir de point de départ pour l'analyse de cette discipline, tout comme il peut servir de point de départ en philosophie de la mesure. Pour ce faire, je reformule d'abord le problème de la coordination à deux reprises : une première fois de sorte à révéler toute sa portée, en faisant abstraction du type de pratique auquel il réfère (mesure, classification, etc.); une seconde fois afin de l'arrimer spécifiquement aux pratiques classificatoires.

Sans référence à la mesure, et de manière abstraite, le problème peut être reformulé comme suit : considérant une théorie T , formée des termes théoriques $t_1, t_2, t_3...t_n$, ainsi qu'un phénomène P , comment savoir si T nous renseigne sur P ? Il faut coordonner les termes constitutifs de T avec des aspects de P . Pour ce faire, il est nécessaire d'établir des procédures. C'est au moment de leur établissement que la circularité problématique s'incarne. Alors que Chang considère que le problème de la coordination est une forme handicapante de la charge théorique des observations, la formulation plus générale suggère plutôt qu'il s'agit du fait même de la charge théorique des observations en identifiant les procédures comme médiums de la charge théorique en question.

Afin de tenir compte des spécificités des pratiques classificatoires, il faut reformuler une seconde fois le problème : considérant une théorie T , composée d'objets $x_1, x_2, x_3...x_n$ ainsi que de classes d'objets $C_1, C_2, C_3... C_n$ en relation au sein d'un système S , et un phénomène P , comment

savoir si T nous renseigne sur P ? Deux réponses complémentaires peuvent alors être offertes: 1) il est possible d'établir l'appartenance de P ou de l'une de ses composantes à $C_1, C_2, C_3...$ ou C_n ; 2) il est aussi possible d'identifier les relations entre P et d'autres phénomènes (ou entre différents aspects de P) afin de savoir si le système S est bien réalisé. Au moins deux types de connaissances peuvent être ainsi empiriquement construites : d'une part, des connaissances portant sur une entité, en présumant que le fait qu'elle appartienne à une classe nous informe à son sujet; d'autre part, des connaissances au sujet des relations entre classes d'objets. Pour ancrer ces connaissances dans l'expérience, il est nécessaire d'établir des procédures d'observation (pour les relations formant le système ou l'appartenance des objets à une classe). C'est au moment de leur mise en place que la circularité problématique est incarnée.

Il convient de démontrer que cette formulation du problème de la coordination peut elle-même être coordonnée avec des phénomènes, soit des pratiques de classification scientifique (cette affirmation implique effectivement que le problème de la coordination s'incarne aussi dans les descriptions qualitatives de la philosophie des sciences, mais je laisse cette avenue ouverte pour de futures recherches). La cladistique, telle que formulée par Hennig (1965, 1966, 1981), son fondateur, peut servir d'exemple. Malgré les transformations que la cladistique et l'utilisation de ce terme ont subies au travers du temps (voir Quinn 2017), les travaux de Hennig demeurent un jalon important de la discipline qui permet d'explicitier le lien de justification unissant théorie de l'évolution et pratiques phylogénétiques avec une grande clarté grâce aux scrupules méthodologiques de l'auteur en question.

Avant de détailler cet exemple d'incarnation du problème de la coordination au sein de pratiques classificatoires, je tiens cependant à préciser que les travaux de Hennig ont fait l'objet de nombreuses études historiques (par exemple, Rieppel 2016) et philosophiques (par exemple, Hull 1970, 2001). Le présent chapitre doit être reçu comme un début de réflexion entourant le problème de la coordination et son utilité pour l'analyse de pratiques phylogénétiques et non comme une contribution à la compréhension des dynamiques scientifiques à l'époque de Hennig. De plus, la cladistique à laquelle je réfère est exclusivement celle que l'on peut lire chez Hennig. Je ne présume aucunement que mon analyse peut s'appliquer aux travaux, hautement hétérogènes, de toutes les personnes s'associant ou ayant été associées à la cladistique; plus

spécifiquement, les défenseurs et défenseuses des méthodes d'inférence phylogénétique basées sur la parcimonie, qui reçoivent et s'autoattribuent parfois l'étiquette cladiste, sont exclu·e·s du propos de la section suivante (Quinn 2017). Afin d'accomplir cette tâche restreinte, je vais d'abord résumer en quoi consiste la cladistique telle que formulée par Hennig, puis expliciter le lien intime qu'elle entretient avec la théorie darwinienne de l'évolution, et du même coup, mettre en lumière le fait que la cladistique hennigienne est soumise au problème de la coordination.

1.4 Le problème de la coordination et les pratiques classificatoires : l'exemple de la cladistique hennigienne

1.4.1 Le fonctionnement inférentiel de la cladistique hennigienne

La cladistique, telle qu'articulée par Hennig, est un ensemble de pratiques classificatoires qui décrit la biodiversité sur la base d'espèces d'organismes et en fonction du degré de parenté relatif entre ces espèces, représentant le tout par des cladogrammes. Un cladogramme est un diagramme arborescent (un dendrogramme) au sein duquel l'apparition de traits dérivés, des synapomorphies (intimement liées à la notion d'homologie; voir prochaine section et Patterson 1982), est représentée par des embranchements successifs. Les synapomorphies lient deux espèces, dès lors dites sœurs, au sein de clades monophylétiques, c'est-à-dire qu'il s'agit, au sens de Hennig, de groupes composés exclusivement d'une espèce ancestrale et de tous ses descendants. Par l'identification des apparitions de traits successives qui composent le cladogramme (chaque embranchement représente une telle apparition), la cladistique informe au sujet des espèces par l'entremise de leurs liens de parenté relatifs. Ce lien de parenté est dit relatif, puisqu'il n'offre qu'une mesure de la parenté qui est comparative : par exemple, l'espèce A est généalogiquement plus proche de l'espèce B que de l'espèce C (Hull 1970; Panchen 1992; Wiley 1981; Wiley et Lieberman 2011).

Ces degrés de parenté peuvent ensuite être mobilisés pour construire des systèmes taxonomiques dans lesquels les espèces les plus proches sont réunies au sein de clades, qui sont à leur tour ordonnés selon leurs liens de parenté relatifs, et ainsi de suite, de sorte que la cladistique permette la construction de systèmes hiérarchiques intégraux. Pour Hennig, un lien

d'isomorphie existe entre phylogénie et classification, une affirmation contestée à la fois par d'autres approches cladistes et les approches phylogénétiques, mais non cladistes (Hennig 1975; Hull 1970; Mayr 1969a, 1974, 1981; Panchen 1992; Quinn 2017; Wiley 1981)⁴.

C'est donc dire que la notion de synapomorphie est au cœur du processus inférentiel de la cladistique, puisque c'est à partir de l'identification des synapomorphies qu'il devient possible d'inférer les liens de parenté entre espèces. L'un des hauts faits d'armes de Hennig a été de détailler les outils disponibles pour analyser les traits morphologiques de manière à pouvoir retracer la phylogénie, cette dernière étant pensée simplement comme étant la succession d'embranchements qui a donné naissance aux espèces étudiées. Les synapomorphies sont des traits dérivés, qui sont partagés par certaines espèces, mais surtout qui distinguent les espèces sœurs des autres espèces avec lesquelles elles partagent un ancêtre commun plus lointain. Autrement dit, la synapomorphie apparaît chez le plus proche ancêtre commun de deux espèces-sœurs, un ancêtre qu'elles ne partagent pas avec d'autres espèces, et leur est ainsi léguée. Les synapomorphies sont à distinguer des autapomorphies (traits distinctifs d'une espèce), des homoplasies (traits similaires chez deux espèces, mais ayant évolué séparément), des plésiomorphies (traits ancestraux plus lointains et conservés, appelés simplésiomorphies lorsqu'ils sont présents au sein d'espèces issues de groupes monophylétiques différents), etc. Une fois les synapomorphies identifiées, Hennig suggère de les utiliser pour reconstruire la phylogénie des groupes étudiés grâce à l'information très spécifique qu'elles fournissent sur l'histoire évolutive des espèces. Cela dit, ces inférences (qui nous font passer d'analyse de traits à des

⁴ Il a été argumenté que la cladistique classique est incapable de reconstruire l'histoire des espèces *per se*. Elle ne serait qu'en mesure de fournir un ordre d'apparition des états des traits observés, avec lequel une pluralité d'histoires évolutives des espèces est compatible (Panchen 1992, 56-58). Cette précision, fort importante, est ici mise de côté puisque les critiques allant en ce sens ne s'appliquent pas aux pratiques phylogénétiques contemporaines. Je continue donc de centrer mon propos sur le problème de la coordination et son incarnation phylogénétique.

Il convient aussi de rappeler que les biologistes font une distinction entre les méthodes phylogénétiques permettant de reconstruire des hypothèses relatives à l'histoire des espèces, d'un côté, et la classification des organismes, de l'autre, même lorsque cette classification est réalisée sur la base des hypothèses phylogénétiques. Bien que je reconnaisse l'importance de cette distinction, toute pratique qui vise à ordonner des catégories d'objets, par exemple sur la base de liens phylogénétiques, est incluse dans cette thèse dans l'ensemble auquel je réfère par l'expression « pratiques classificatoires ». L'hétérogénéité des pratiques classificatoires doit cependant être prise en compte lorsque cela influe l'analyse, comme en témoigne la formulation du problème de la coordination que j'ai donnée dans la section précédente.

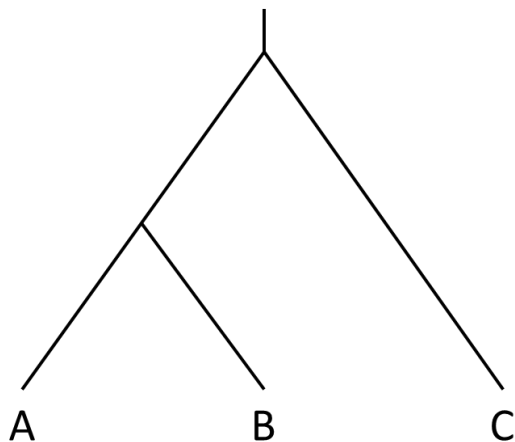
reconstructions phylogénétiques) mobilisant les synapomorphies requièrent des informations externes pour être réalisées, par exemple lorsque des signaux phylogénétiques contradictoires sont observables.

Illustrons la chose par un exemple (issu de Panchen 1992, 153-155) qui simplifie l'étude des groupes concernés, mais qui a le mérite d'être éloquent. Considérons trois groupes, soit A, les oiseaux (*Aves*); B, les crocodiliens (*Crocodylia*); et C, l'ensemble des autres reptiles vivants. A, B et C forment un taxon supérieur, soit les *Sauropsidae*. La question qui se pose alors est la suivante : quels groupes sont des groupes frères? L'identification de synapomorphies présentes chez deux des trois groupes, c'est-à-dire de traits qu'ils auraient hérités d'un ancêtre commun à ces deux groupes et qu'ils ne partagent pas avec le troisième, permettrait de répondre à la question (ces deux groupes seraient des groupes frères).

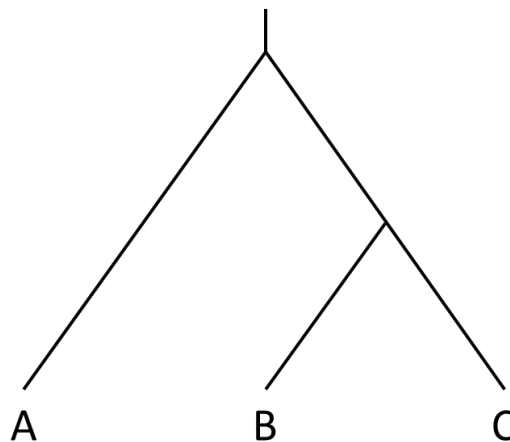
La séparation complète des ventricules semble être une synapomorphie reliant crocodiliens et oiseaux au sein d'un clade monophylétique. Au contraire, l'absence de l'arc aortique du côté gauche chez les oiseaux, un trait présent chez les crocodiliens et les autres reptiles, invite à grouper B et C ensemble plutôt que A et B. La différence est phylogénétiquement significative puisque les deux hypothèses (illustrées par la Figure 1.1) suggèrent des histoires évolutives du groupe ABC drastiquement différentes, ayant notamment une grande influence sur l'importante question « les oiseaux sont-ils des dinosaures? ». Pour régler le débat, il faut savoir lequel des deux traits est effectivement une synapomorphie, comme définie par Hennig, et lequel offre un signal phylogénétique qui prête à confusion.

Figure 1.1 – Deux hypothèses phylogénétiques pour le groupe A (*Aves*), B (*Crocodylia*) et C (les autres reptiles)

1.1.a. La séparation complète des ventricules est une synapomorphie reliant A et B



1.1.b. La présence de l'arc aortique à gauche serait une synapomorphie reliant B et C



Cette figure est une représentation de deux hypothèses différentes au sujet de l'histoire phylogénétique des groupes A, B et C. Chaque embranchement représente un événement de spéciation lors duquel un groupe ancestral donne naissance à deux nouveaux groupes. L'hypothèse 1.1.a suggère que les oiseaux et les crocodiliens sont des groupes frères (hypothèse acceptée à ce jour), alors que 1.1.b suggère que les crocodiliens et les autres reptiles sont des groupes frères. D'un point de vue phylogénétique, la différence principale entre ces deux hypothèses est la présence d'un ancêtre commun à A et B mais pas à C (1.1.a) ou, à l'inverse, d'un ancêtre commun à B et C mais pas à A (1.2.b).

L'une des façons de réaliser de telles discriminations, développée par Hennig, est de mobiliser un groupe extérieur à l'ensemble pour lequel on tente de reconstruire l'histoire évolutive afin d'accroître l'information disponible (la mobilisation de groupes externes est encore aujourd'hui utilisée de manière courante pour reconstruire des phylogénies). Dans le cas qui nous intéresse, les mammifères (formant les *Theropsidae*), groupe frère des *Sauropsidae*, peut être mobilisé. Les mammifères, comme les groupes B et C, possèdent une arche aortique à gauche. Celle-ci doit alors être considérée comme une symplesiomorphie, un trait hérité d'un ancêtre plus reculé, qui a été conservé dans tous ces taxons, à l'exception des oiseaux. Il ne s'agit *pas* d'un trait dérivé propre à B et C, de sorte que l'hypothèse la mieux soutenue par l'observation est celle d'une évolution commune plus récente des groupes A et B, les oiseaux et les crocodiliens.

Cette logique est rendue possible par l'adhésion des praticiens et praticiennes au principe de parcimonie, selon lequel la meilleure hypothèse phylogénétique est celle nécessitant le moins de modifications d'états de caractères possible. En effet, pour faire de l'arche aortique à gauche

une synapomorphie reliant les groupes B et C, il faudrait que ce trait soit apparu indépendamment deux fois plutôt qu'une, chez les mammifères et chez l'ancêtre commun à B et C. La séparation complète des ventricules, pour être considérée comme étant une synapomorphie reliant A et B, ne requiert aucun doublon du genre. Il s'agit donc de l'hypothèse la plus parcimonieuse, soit celle à prioriser. Ce type de raisonnement est caractéristique de la cladistique hennigienne, qui peut se résumer par la phrase suivante : « A shares a common ancestor with B not shared by C » (Hennig 1966, cité dans Panchen 1992, 153).

1.5 Le rapport de justification mutuelle unissant la cladistique hennigienne et la théorie l'évolution par voie de sélection naturelle

Maintenant que le fonctionnement inférentiel typique de la cladistique hennigienne a été illustré, au moins succinctement, il convient de se pencher sur le lien qu'elle entretient avec la théorie de l'évolution par voie de sélection naturelle. Pour le dire simplement, l'opération même de reconstruire le passé phylogénétique implique minimalement que l'on considère qu'une telle chose existe. Il faut donc rejeter toute forme de fixisme. De plus, et c'est ce sur quoi il convient d'insister, la théorie de l'évolution par voie de sélection naturelle acquiert son contenu empirique notamment par l'entremise des travaux phylogénétiques qui viennent confirmer ses prédictions quant aux relations entre espèces. Il convient de présenter cette relation de soutien réciproque en détail.

La théorie de l'évolution par voie de sélection naturelle⁵ a comme *explanandum* l'état de la biodiversité sur Terre. Cette théorie permettait en effet à Darwin d'expliquer pourquoi certaines espèces se ressemblent plus entre elles qu'elles ne ressemblent à d'autres et pourquoi

⁵ Je réfère ici et plus loin à la théorie de l'évolution par voie de sélection naturelle que je qualifie parfois de darwinienne afin de simplifier le cas d'étude. Hennig a en effet écrit pendant la deuxième moitié du 20^e siècle, de sorte que son adhésion à la théorie de l'évolution renvoie à la Synthèse moderne. Cependant, les précisions relatives aux dynamiques des populations et à la qualification de la variation que la Synthèse moderne a su mettre de l'avant n'influent pas mon analyse de la cladistique hennigienne. Je réfère donc, de manière générale, à la théorie de l'évolution *darwinienne* afin de souligner la stabilité des principes théoriques plus généraux qui l'animent depuis sa première formulation.

les êtres vivants sont aussi bien adaptés à leur environnement (incluant d'autres êtres vivants). Selon lui, la théorie de la descendance avec modification permettait aussi de comprendre pourquoi, dans certains cas, il est si difficile de distinguer les groupes d'organismes qui forment une espèce de ceux qui forment des variétés (un rang taxonomique réputé inférieur à celui d'espèce). En d'autres termes, Darwin a construit une théorie qui explique, entre autres, les systèmes taxonomiques auxquels il avait accès.

Cette théorie darwinienne, dans sa version traditionnelle, se résume comme suit. Les organismes les mieux adaptés à leur environnement (cet avantage peut être infime et induit par une variation individuelle qui pourrait sembler insignifiante) tendent à laisser plus de descendants que leurs voisins. Puisque la présence d'hérédité est indéniable dans le monde vivant, les descendants des organismes avantagés seront probablement avantagés eux aussi, en plus de pouvoir bénéficier (avec un peu de chance) de nouvelles variations elles aussi plus bénéfiques. Certaines lignées d'organismes prolifèrent ainsi, alors que d'autres, au contraire, tendent vers l'extinction. Ce tri, imputable aux conditions écologiques qui déterminent quels organismes ont du succès et lesquels n'en ont point, est ce que Darwin appelle la sélection naturelle. L'évolution par voie de sélection naturelle est le résultat de ce tri appliqué sur de nombreuses générations, un tri qui favorise, selon Darwin, les lignées qui se renouvellent constamment, de sorte que l'apparition de nouvelles espèces (définies par Darwin comme étant des communautés de descendance), à partir de variations individuelles se généralisant (se fixant dans une population, pour utiliser un jargon contemporain) progressivement (en passant par le stade de la variété), est inévitable.

Sous sa forme abstraite et encore plus schématisée, la théorie de l'évolution par voie de sélection naturelle affirme qu'une population d'entités se transforme (ou se stabilise) au travers du temps sous l'influence des pressions environnementales si cette population est composée d'individus qui présentent des variations individuelles, des différences de fitness (mesure de la capacité à survivre et se reproduire) découlant de ces variations et de l'hérédité (Lewontin 1970; Godfrey-Smith 2009; Papale 2021). Si ces trois conditions sont remplies, l'accumulation de mutations favorables chez les individus au fil des générations successives peut engendrer une transformation de la population, en ce sens que les individus qui la composent au temps $t+1$ sont

différents de ceux qui la composaient au temps t . De plus, si le contexte écologique demeure stable et que la population est assez grande, on peut s'attendre à ce que les mutations s'accumulent de sorte à engendrer des adaptations, c'est-à-dire des traits avantageux pour les individus qui les portent dans le contexte écologique en question et dont la présence au sein de la population est due à cette utilité inscrite dans leur passé évolutif.

D'un point de vue épistémique, l'un des défis principaux pour les biologistes de l'évolution est de fournir un contenu empirique à cette théorie. Démontrer que l'évolution d'une population ou d'une lignée peut être guidée par la sélection naturelle est déjà un défi de taille en soi. Démontrer que la sélection naturelle est le vecteur principal de l'évolution sur Terre, comme les darwiniens et darwiniennes⁶ du 19^e siècle et les tenants et tenantes de la synthèse moderne l'ont défendu par la suite, est un objectif encore plus imposant. Cet objectif, de nombreuses disciplines (géologie, biogéographie, génétique des populations, taxonomie, phylogénétique, statistique et mathématique plus généralement, informatique, génétique, microbiologie, biochimie, etc.) ont contribué et contribuent encore aujourd'hui à le remplir avec brio. La phylogénétique, dans laquelle s'inscrit la cladistique hennigienne, occupe une place de choix au sein de ce consortium.

La logique soutenant l'hypothèse darwinienne par l'entremise de la phylogénétique est la suivante : si le système de relations entre organismes et espèces que reconstruisent les inférences phylogénétiques concorde avec les prédictions qu'on peut faire sur la base de la théorie darwinienne, alors celle-ci semble bien justifiée (Penny, Foulds et Hendy 1982). L'une des prédictions faites par la théorie en question est que d'une seule lignée peut naître un grand nombre d'espèces. Toutes ces espèces forment ensemble un groupe monophylétique. Chaque taxon supérieur (de rang supérieur à celui d'espèce) monophylétique avéré peut ainsi être considéré comme étant une preuve de la validité de l'hypothèse darwinienne qui postule

⁶ Tout au cours de l'ouvrage, j'entends utiliser la rédaction épïcène même lorsque la situation suggère que seulement des hommes pourraient être concernés. Cette posture vise à confronter l'invisibilisation des femmes, trop souvent caractéristique de l'historiographie des sciences (Flores Espinola 2013). Par exemple, il semble facile de croire qu'au 19^e siècle, il n'y avait que des darwiniens dignes de mention au sein de l'académie, une affirmation contredite par les travaux et la contribution importante, quoique souvent négligée, de Blackwell (Blackwell 1875; Hrdy 1999). L'ampleur de la tâche que cela impliquerait rend impossible la justification de chacune des occurrences de rédaction épïcène dans cette thèse, mais je tenais à présenter au moins celle-ci.

l'existence et explique l'émergence (évolution par accumulation de variations individuelles et divergences des lignées) de ces groupes monophylétiques.

Certains phylogénéticiens et certaines phylogénéticiennes expriment la chose clairement, comme en témoigne la citation suivante, issue de l'introduction d'un ouvrage de référence :

Evolutionary theory predicts that monophyletic groups and only such groups emerge from various evolutionary processes termed *speciation*. They are composed of a common ancestral species and all of that species' descendants. [...] Such groupings are sought because evolutionary theory predicts their existence. [...] Thus, all truly monophyletic groups have the property of being composed of species, or higher taxa, who have exclusive, or unique, genealogical descent from a founder species. Each higher taxon we hypothesize to be monophyletic stands as a singular confirmation of macroevolutionary theory because macroevolutionary theory predicts that such groupings should exist. (Wiley et Lieberman 2011, 18)

Cette prédiction est intimement liée à l'idée selon laquelle les patrons phylogénétiques issus de l'évolution sont nécessairement arborescents. En effet, pour Darwin, à l'époque, et les biologistes d'hier et d'aujourd'hui, la prédominance des groupes monophylétiques au sein des patrons généalogiques s'explique notamment par la priorité des trajectoires divergentes au sein des dynamiques évolutives : les espèces divergent les unes des autres et, à l'inverse, la convergence de lignées, qui donne lieu à une réticulation généalogique, est rare et négligeable. Or, si les différentes lignées ne peuvent que diverger les unes des autres, il en résulte un patron arborescent, comme en témoigne fameusement le seul graphique de *L'Origine* (Figure 1.2), au sein duquel il est possible d'identifier des groupes contenant tous et exclusivement les descendants d'un ancêtre commun, des groupes monophylétiques.

reconstructions phylogénétiques) comme étant des hypothèses, *en pratique*, ce sont des résultats aussi. Ces « hypothèses » peuvent donc bel et bien confirmer des théories évolutives selon la logique de la discipline.

Les questions qui s'imposent alors sont du même ordre que celles soulignées par la philosophie de la mesure : quelles procédures permettent de vérifier si le patron arborescent et les groupes monophylétiques existent bien dans le monde, au-delà de la théorie? Comment peut-on savoir si la théorie explique adéquatement leur présence? Comment peut-on justifier l'utilisation de ces procédures? La cladistique et la phylogénétique en général offrent des réponses à la première question. De ces réponses, une analyse peut être faite pour appréhender les deux autres questions. Maintenant que le lien justificatif entre cladistique et théorie de l'évolution est détaillé, il est possible de bien circonscrire le problème de la coordination au sein de cette structure épistémique.

1.5.1 La structure épistémique de la cladistique hennigienne et le problème de la coordination

Du point de vue de la justification, le défi central auquel fait face la phylogénétique est très certainement l'invisibilité des liens de parenté entre les espèces, des liens qui s'inscrivent dans un passé auquel les sens humains ne peuvent accéder. Il faut donc inférer ces relations à partir d'autres types de relations elles-mêmes observables, de sorte que la vérification d'hypothèses (existence de groupes monophylétiques, patron arborescent des systèmes phylogénétiques, etc.) dépend de procédures d'inférence reliant généalogie et quelque chose d'autre, nommément la similarité (ou dissimilarité) morphologique ou génétique. Sur le plan de la structure justificative, l'observation de la phylogénie est de ce fait comparable à l'observation de la température qui relève de sa corrélation avec pression et volume.

L'importance de la contribution de Hennig à la phylogénétique provient de sa méthode simple pour inférer des relations phylogénétiques entre espèces à partir de la similarité, une méthode décrite plus tôt. Or, ces procédures (identification de similarité, analyse des signaux phylogénétiques offerts par cette similarité, formation des clades sur la base de cette information) se font au prix d'une série de prémisses qui doivent être acceptées de fait pour

réaliser des inférences phylogénétiques comme prescrites par la cladistique. J'ai extrait sept de ces croyances aprioriques du texte de Hennig de 1965, toutes exprimées explicitement par l'auteur : 1- la théorie de l'évolution darwinienne est valide; 2- les espèces sont définies comme étant des communautés reproductives; 3- la similarité morphologique reliant deux individus peut être corrélée aux relations de parenté; 4- les événements de spéciation sont dichotomiques; 5- la reproduction sexuée est la norme; 6- le degré de parenté traduit par la classification est relatif; 7- l'hypothèse la plus simple doit être priorisée. Cette liste n'est pas exhaustive.

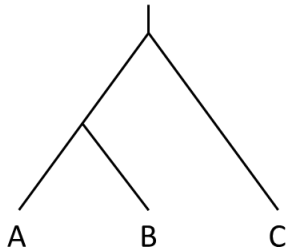
C'est seulement par l'acceptation de tels principes qu'il devient possible d'observer des groupes d'organismes et d'inférer les liens phylogénétiques qui les uniraient. Reprenons l'exemple détaillé précédemment, qui implique les groupes A (oiseaux), B (crocodiliens) et C (autres reptiles), formant un taxon monophylétique. Dans une perspective cladiste (Hennig 1965; Panchen 1992), deux hypothèses phylogénétiques sont disponibles pour ces trois groupes en fonction des traits observés (présence/absence de l'arche aortique du côté gauche, séparation complète ou non des ventricules) : A et B sont des groupes frères ou B et C le sont.

Cela dit, beaucoup d'autres hypothèses pourraient être formulées si la créativité des chercheurs et des chercheuses pouvait être exemptée des contraintes imposées par les prémisses mentionnées précédemment. Les trois groupes pourraient être frères, c'est-à-dire qu'une espèce ancestrale leur aurait donné naissance simultanément; les trois espèces auraient alors le même degré de parenté les unes avec les autres. Rien ne nous empêche non plus de postuler que C et A soient des groupes frères, et que les ressemblances entre A et B ainsi qu'entre C et B sont des simplésiomorphies plutôt que des synapomorphies, voire des homoplasies. D'autres hypothèses seraient encore possibles pour les trois groupes A, B et C (réticulation, fixisme et donc indépendance phylogénétique des espèces, apparition des espèces au sein d'une séquence linéaire, la longueur des branches pourrait être spécifiée et considérée importante, etc.); la Figure 1.3 exemplifie la diversité des systèmes de parenté entre groupes taxonomiques qui peuvent être postulés lorsque les contraintes théoriques sont relâchées) et le nombre d'hypothèses possibles

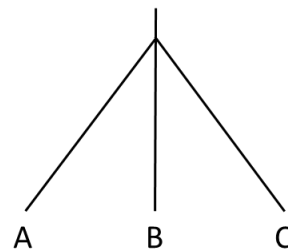
augmente très rapidement⁷ par rapport au nombre de groupes pour lesquels on cherche à retracer une phylogénie.

Figure 1.3 – Exemples d'hypothèses phylogénétiques variées pour le groupe A, B, C avec des contraintes théoriques relâchées

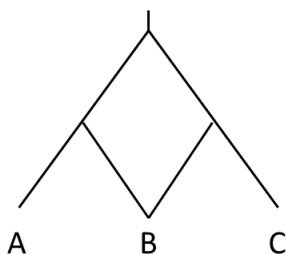
1.3.a. Hypothèse standard avec spéciation dichotomique



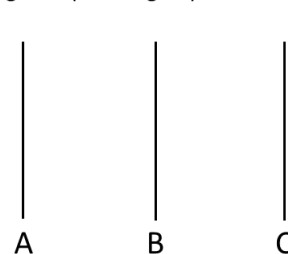
1.3.b. Hypothèse avec spéciation trichotomique permise



1.3.c. Hypothèse avec réticulation



1.3.d. Hypothèse avec indépendance phylogénétique des groupes taxonomiques



Dans cette figure, quatre exemples d'hypothèses phylogénétiques (portant sur les liens de parenté entre groupes taxonomiques) variées sont représentées afin d'illustrer l'immensité du champ des possibles qui émerge lorsque l'on rejette les contraintes théoriques de base de la phylogénétique (en tant que discipline). Les hypothèses 1.3.b, 1.3.c et 1.3.d sont généralement considérées comme étant irrecevables puisqu'elles ne se conforment pas au modèle arborescent. L'hypothèse 1.3.c, qui présente de la réticulation phylogénétique (B serait issu d'interactions entre A et C), illustre le type d'événement phylogénétique que les approches réseaux mettent en lumière. L'hypothèse 1.3.d est incompatible avec certaines des prémisses les plus fondamentales du paradigme évolutif, puisqu'elle implique la création indépendante des unités taxonomiques.

De ce fait, la puissance calculatoire nécessaire pour reconstruire (observer) des arbres phylogénétiques et le passé évolutif des espèces est immense. Sans la standardisation de procédures qui restreignent le champ des possibles (ainsi que les heuristiques mobilisées de nos jours), de telles inférences resteraient de simples ambitions. Cette standardisation se fait notamment par l'entremise des prémisses mentionnées plus tôt. Celles-ci contraignent les hypothèses et déterminent ainsi les procédures (inférentielles) nécessaires pour l'observation des

⁷ La formule qui traduit le nombre d'arbres possibles est $(2n - 3)! / 2^{n-2}(n - 2)!$ pour des arbres enracinés qui adoptent la spéciation dichotomique, où n est le nombre d'unités taxonomiques concernées.

liens phylogénétiques entre les espèces. Par exemple, le principe selon lequel la spéciation est toujours dichotomique permet de rejeter l'hypothèse selon laquelle les trois groupes A, B et C seraient tous aussi proches phylogénétiquement les uns des autres (prémisse 4; pour Hennig, la distance phylogénétique entre deux groupes taxonomiques réfère exclusivement au nombre d'embranchements qui les sépare). Puisque chaque embranchement n'engendre que deux lignées (qui peuvent éventuellement incarner des taxons supérieurs par la mise en place d'embranchements successifs), alors deux espèces ont forcément émergé d'un embranchement plus récent que celui duquel la troisième est issue (voir le contraste entre les Figures 1.3.a et 1.3.b)⁸. De surcroît, toute inférence serait impossible si la croyance apriorique selon laquelle la phylogénie peut être inférée sur la base de la similarité morphologique (et plus tard, génétique) manquait à l'appel (prémisse 3).

En d'autres termes, la cladistique offre un soutien empirique pour la théorie darwinienne, mais infère la phylogénie avec l'aide d'éléments théoriques issus de ladite théorie. Il s'agit incontestablement d'une incarnation du problème de la coordination. Regardons de plus près comment les principes 3 et 4 cristallisent le va-et-vient entre théorie et procédures classificatoires en mettant l'accent sur leur rôle au sein de la théorie darwinienne. Le même travail pourrait être effectué à partir des autres prémisses (sauf peut-être pour la première, qui est trop franchement darwinienne pour donner lieu à une analyse détaillée).

La théorie darwinienne de l'évolution est ancrée dans l'empirie par la corrélation entre descendance commune et traits similaires dont la sélection naturelle forge le dynamisme (prémisse 3); en effet, l'hérédité est largement reconnue comme étant une condition nécessaire à la réalisation de l'évolution par voie de sélection naturelle (Darwin 1859; Lewontin 1970; Godfrey-Smith 2009). Or, l'hérédité réfère à l'idée selon laquelle les liens de parenté engendrent des liens de similarité, puisque des mécanismes causaux garantissent une certaine transmission

⁸ Il est à noter que dans le cadre de reconstructions phylogénétiques non enracinées (des arbres au sein desquels l'ancêtre commun de toutes les espèces est absent), pour des arbres à trois espèces, la spéciation trichotomique est permise et même nécessaire. Or, dès que l'on passe à des arbres non enracinés d'au moins quatre espèces, la spéciation dichotomique s'impose (toujours dans une perspective traditionnelle). De surcroît, les arbres enracinés sont quant à eux irrémédiablement dichotomiques.

d'états de caractères d'un parent à ses rejetons. Doublée de l'apparition de nouveautés (typiquement considérées comme résultant de mutations) et de l'action de la sélection naturelle qui élague les formes mal adaptées à leur milieu, cette transmission plus ou moins fidèle de traits permet l'accumulation de modifications bénéfiques (dans un contexte donné) et éventuellement l'apparition de ce qu'on appelle des adaptations. En d'autres termes, la théorie stipule que les adaptations, ces traits dont elle cherche à expliquer l'apparition, émergent parce qu'il y a des liens de descendance avec hérédité entre les individus au sein d'une population. À la lumière de ce discours, il est facile de saisir pourquoi les phylogénéticiens et phylogénéticiennes considèrent qu'il est acceptable, épistémiquement parlant, d'inférer la parenté sur la base de la similarité : les deux sont, selon la théorie, inextricables⁹. Cette pratique inférentielle semble ainsi justifiée, mais il n'en demeure pas moins que la cladistique, et toute pratique phylogénétique, mobilise des éléments de la théorie qu'elle contribue à corroborer empiriquement, générant ce qui semble être une inquiétante circularité.

Tout comme la prémisse 3, la quatrième est elle aussi inscrite au cœur de la théorie darwinienne. Comme déjà mentionné, le principe 4 implique que le seul type d'événement menant à la naissance d'une espèce que la cladistique entérine est la spéciation dichotomique. La convergence, l'hybridation et tout autre type d'événement de spéciation qui créerait de la réticulation sont *de facto* exclus des cladogrammes et des hypothèses cladistiques. Pour Hennig, ce point est tellement crucial qu'il en vient à déterminer la notion même de relation phylogénétique, c'est-à-dire la cible de toute étude cladistique : « Thus, all species (= reproductive communities) which exist together at a given time, e.g., the present, have originated by the splitting of older homogeneous reproductive communities. On this fact is based the definition of the concept, "phylogenetic relationship" » (Hennig 1965, 97). Cette priorisation de la division dichotomique des espèces ancestrales comme événement évolutif mène à la formation de systèmes phylogénétiques traversés par une constante, soit la divergence.

⁹ Dans la section portant sur la notion d'homologie ainsi qu'au chapitre 5, je donne un aperçu de la richesse théorique inscrite dans le terme « similarité » par les phylogénéticiens et phylogénéticiennes. La transition vers la similarité moléculaire, de laquelle je traite au chapitre 4, n'a fait qu'accroître la complexité sémantique de ce terme.

Dès la première formulation de la théorie de l'évolution par voie de sélection naturelle, Darwin fait de ce qu'il appelle le principe de divergence la principale caractéristique des patrons phylogénétiques prédits par l'action de la sélection naturelle :

Natural selection, as has just been remarked, leads to divergence of character and to much extinction of the less improved and intermediate forms of life. On these principles, the nature of the affinities, and the generally well-defined distinctions between the innumerable organic beings in each class throughout the world, may be explained. (Darwin 1859, 97)

La spéciation dichotomique, doublée du principe de divergence, est donc bien inscrite au cœur de la théorie que la cladistique hennigienne et ses modèles arborescents cherchent à ancrer empiriquement.

Toutes ces avenues d'analyse convergent pour mettre en lumière une structure justificative circulaire unissant théorie de l'évolution et pratiques classificatoires. Cette apparente circularité, Hennig l'a reconnue lui-même, bien que ses conséquences épistémiques aient été systématiquement sous-estimées : « Since the advent of the theory of evolution, one of the tasks of biology has been to investigate the phylogenetic relationship between species » (Hennig 1965, 97). Un peu plus loin, dans le même article, il ajoute la chose suivante : « Biological systematics can no more do without a theoretical foundation for its work than can any other science » (Hennig, 1965, 115). La situation peut être reformulée en termes compatibles avec le problème de la coordination : les relations phylogénétiques entre organismes et groupes d'organismes sont inobservables par la perception humaine seule; il faut les inférer à partir de caractéristiques observables, comme la similarité (en postulant que des méthodes d'observation pour la similarité aient été préalablement acceptées). Ces inférences requièrent des procédures, comme la cladistique hennigienne imbibée de darwinisme, qui permet d'inférer la phylogénie notamment à partir d'une analyse de la similarité. Pour vérifier de manière absolue la validité de la procédure en question, il faudrait pouvoir observer la phylogénie indépendamment de la similarité, mais l'observation de la phylogénie est justement ce que l'on cherche à rendre possible (notamment, mais non exclusivement, pour fournir un contenu empirique à la théorie de l'évolution par voie de sélection naturelle). Le principe de base de l'empirisme est que l'expérience doit être mobilisée

pour soutenir toute théorie. Le problème de la coordination souligne que cette expérience requiert une charge théorique. Cette situation est bien présente en phylogénétique.

1.5.2 Un autre aspect du problème de la coordination : la formation de catégories

Je n'ai pour l'instant exposé qu'un seul aspect du problème de la coordination tel qu'il s'applique aux pratiques classificatoires. Dans la section précédente, lorsque j'ai reformulé le problème de la coordination pour qu'il s'applique au-delà des pratiques métrologiques, j'ai bien précisé qu'au moins deux types de connaissances peuvent être construites dans une perspective classificatoire : des connaissances portant sur une entité, en présupposant que le fait qu'elle appartienne à une classe nous informe à son sujet, et des connaissances au sujet des relations entre classes d'objets. Dans ce qui précède, j'ai montré comment le deuxième aspect est bien présent en phylogénétique, mais je tiens à préciser, au moins rapidement, que le premier l'est aussi.

La phylogénétique s'est traditionnellement intéressée à l'espèce comme unité de classification. Malgré la diversité de définitions de la notion d'espèce disponibles en biologie, la phylogénétique mobilise le concept d'espèce biologique fameusement défendu par Mayr (de Queiroz 2007; Mayr 1969b), mais aussi adopté par Hennig (1965). En ce sens, une espèce est une communauté de descendance, c'est-à-dire qu'elle rassemble tous et seulement les descendants d'un ancêtre commun; cette conception est d'ailleurs en accord avec la définition que Darwin offre au deuxième chapitre de *L'Origine des espèces*.

La question qui se pose ici consiste à savoir comment identifier l'appartenance d'un individu à une espèce donnée, considérant que le passé généalogique de cet individu est de fait inobservable (directement). Exactement comme pour les liens phylogénétiques entre espèces, les liens de parenté entre organismes doivent être inférés sur la base de la similarité entre ceux-ci. Ce point, Hennig le reconnaît explicitement lorsqu'il traite des difficultés liées à la classification des organismes au sein d'espèces :

The difficulty within this task rests in the fact that the species, which exist in nature as real phenomena independent of the [humans] who perceive them, are units which are

not morphologically but genetically defined. They are communities of reproduction, not resemblance. *Of course, the morphological resemblance between members of a species is not unimportant for the practice establishment of specific limits.* (Hennig 1965, 101, italiques ajoutés)

Le fait de pouvoir se reposer sur la ressemblance entre organismes pour tracer les limites des espèces tire évidemment sa légitimité de l'acceptation de la puissance de l'hérédité (le semblable engendre le semblable), notion au cœur de l'argument darwinien et au moins aussi vieille que la biologie elle-même.

Or, même si ce lien entre morphologie et appartenance à une espèce est extrêmement bien ancré au sein de la culture scientifique, il n'en demeure pas moins que d'accepter l'hérédité est un prérequis théorique nécessaire à l'observation (par inférence) des liens de parenté entre organismes, et donc de la mise en œuvre du concept d'espèce biologique puisque les liens de parenté eux-mêmes sont inaccessibles à nos sens. La même logique s'applique à toute unité taxonomique opérationnelle (OTU, d'après l'expression anglaise *operational taxonomic unit*) utilisée de nos jours en phylogénétique (la similarité génétique étant aujourd'hui mobilisée pour identifier les frontières d'une OTU).

Dans ce cas-ci comme dans celui détaillé plus longuement et relatif à la reconstruction de systèmes phylogénétiques, la circularité dont il est question est à la fois artéfactuelle et inévitable. D'abord, elle est artéfactuelle en ce sens qu'elle émerge d'une analyse décontextualisée d'une discipline. Dans ce qui précède, j'ai en effet isolé les travaux de Hennig et leur rapport à la théorie darwinienne de leur contexte sociohistorique, ce qui permet de mettre en exergue la structure épistémique circulaire associée au problème de la coordination. Cette simplification, utile pour comprendre le rôle de l'expérience au sein d'une structure justificative comme celle de la phylogénétique, occulte cependant la dimension diachronique des dynamiques scientifiques ainsi que son inclusion au sein d'une culture intellectuelle plus large. Ainsi, le problème de la coordination et ses implications pour l'analyse de pratiques scientifiques doivent être traités dans une optique sociohistorique qui implique un modèle hélicoïdal (Chang 2016). Dans ce qui suit, je démontre que la structure épistémique de la phylogénétique correspond en effet à ce modèle en explorant l'histoire d'un concept central de la phylogénétique, soit l'homologie.

1.6 La structure justificative de la biologie comparative est dynamique, comme illustré par la notion d'homologie

La structure épistémique de la phylogénétique, du moins comme exemplifiée par la cladistique, est caractérisée par un va-et-vient entre théorie et observations, une réciprocity nécessaire à la fois pour établir des procédures d'observations et pour fournir un contenu empirique aux théories scientifiques, comme la théorie de l'évolution par voie de sélection naturelle. Observés de manière anhistorique, ces allers-retours posent un défi sérieux à toute épistémologie empiriste : comment expliquer la validité des connaissances si celles-ci sont le résultat de raisonnements circulaires? Dans le prochain chapitre, j'explore les cadres épistémologiques traditionnels et contemporains qui répondent directement à cette question, en mettant en exergue le problème de la coordination. Cependant, il convient d'abord de compléter l'articulation du rapport entre problème de la coordination et pratiques phylogénétiques. Pour ce faire, il faut dépasser le cas de la cladistique, ou plutôt le remettre dans son contexte sociohistorique, afin de souligner la dimension diachronique du complexe discursif (Tort 1983, p. 524) en question, soit celui auquel je réfère par l'expression « les pratiques phylogénétiques » et qui relève de la mise en relation des espèces ou d'autres groupes taxonomiques en fonction de leurs liens de parenté inférés sur la base de leurs liens de similarité.

Tel qu'on peut l'observer par l'entremise des travaux de Hennig, le rapport entre théorie et pratiques phylogénétiques semble circulaire : on se base sur les principes théoriques issus de la théorie de l'évolution par voie de sélection naturelle pour reconstruire une succession d'embranchements phylogénétiques (d'apparitions de traits desquelles on infère la phylogénie des espèces) qui viennent soutenir la théorie de l'évolution. Précédemment, j'ai donné un exemple concret de cette dynamique : la présence à gauche de l'arche aortique chez les crocodiliens et les autres reptiles devrait être considérée comme une plésiomorphie et non une synapormorphie, ce qui permet de considérer les crocodiliens et les oiseaux comme des groupes frères. Ce raisonnement vient confirmer, selon certains auteurs comme Wiley et Lieberman (voir Wiley et Lieberman 2011, p. 60 de la présente thèse), le patron arborescent ainsi que l'existence

de groupes monophylétiques prédits par la théorie de l'évolution, mais le modèle guidant l'analyse et contraignant le champ des possibles est lui-même darwinien.

Il faut toutefois reconnaître que j'ai triché en présentant cet exemple (pour rendre explicites toutes les étapes du raisonnement). En effet, pour établir le statut de simplésiomorphie de la présence à gauche de l'arche aortique, les cladistes requièrent l'utilisation d'un groupe extérieur (*outgroup*) à l'ensemble A, B et C que l'on cherche à systématiser. Ce groupe permet la polarisation des états de caractères, c'est-à-dire qu'il permet de déterminer lesquels sont des simplésiomorphies et lesquels sont des synapomorphies (Patterson 1982). Une question se pose alors : comment sait-on que le groupe utilisé, en l'occurrence les mammifères, est extérieur au groupe ABC? Comment sait-on qu'il ne forme pas un taxon monophylétique avec, par exemple, B et C, ou juste B? La réponse à ces questions est d'une simplicité importante pour mon propos : on se fie sur des classifications préexistantes qui suggèrent qu'A, B et C sont phylogénétiquement plus proches les uns des autres qu'ils ne le sont des mammifères. Le fait que les mammifères forment un groupe extérieur au groupe A, B et C est bien établi, de sorte qu'il s'agit d'une connaissance sur laquelle des gestes classificatoires subséquents peuvent se reposer en toute confiance.

De ce fait, nous pouvons déjà rejeter la circularité des raisonnements cladistiques puisque ceux-ci s'appuient sur un élément extérieur qui les précède historiquement, soit des résultats classificatoires passés. En d'autres termes, on échange la menace de circularité justificative contre celle de la régression historique infinie, beaucoup plus difficile à nier, puisque les gestes classificatoires passés ont eux aussi été effectués sur la base d'aprioris théoriques qu'ils contribuent à justifier empiriquement. Chaque geste classificatoire est le résultat d'une réciprocité justificative entre procédures d'observation et théorie qui mobilise des résultats passés comme fondement efficace (mais non absolu) aux connaissances classificatoires : « This observation is directed to an apparent problem implicit in the question "Where do out-group comparisons begin?" [...] we agree with Eldredge (1979: 171, footnote) that some progress has been made during the last few centuries, and a practical starting point is to follow the existing classification. » (Watrous et Wheeler 1981, 7).

Comme cette citation en témoigne, l'enjeu dépasse largement l'identification du groupe monophylétique composé des oiseaux et des crocodiliens, soutenue par la connaissance préalable de l'extériorité relative des mammifères. La structure épistémique qui fait progresser les connaissances phylogénétiques s'étire sur « les quelques derniers siècles », de sorte qu'un modèle d'analyse qui cherche à en rendre compte en faisant l'économie de cette dimension historique serait inadéquat. La notion d'homologie permet de souligner l'importance de cette dimension, puisqu'elle est à la fois au cœur de la structure épistémique de la phylogénétique et plus ancienne que la mise en place de la pensée évolutionniste (Panchen 1992, 1994; Rieppel 2015).

1.6.1 L'homologie au sein de la structure épistémique de la phylogénétique

L'homologie est l'un des concepts cruciaux de la biologie comparative et, conséquemment, de la phylogénétique (Bock 1973; Inkpen et Doolittle 2016; Nelson 1994; Panchen 1992; Patterson 1982; Rieppel 2015; Williams 2004). De façon générale, elle réfère à la relation de similarité entre deux traits si ces deux traits sont associés au *même* caractère. Un même caractère peut en effet être présent plusieurs fois dans un organisme (les vertèbres, par exemple) ou dans plusieurs organismes, chaque fois sous différentes formes (chaque forme d'un caractère étant appelée un état de ce caractère). Dans ce qui suit, je m'intéresse principalement à l'homologie morphologique telle que la notion s'est cristallisée dans les années 1970 et 1980, en phylogénétique, sous l'influence de la cladistique hennigienne. En ce sens, l'homologie, alors dite *taxique*, est synonyme de synapomorphie, bien qu'il serait problématique de réduire l'une de ces notions à l'autre en raison de leurs histoires bien distinctes et de leurs utilisations incompatibles dans divers contextes (notamment ailleurs en phylogénétique). Dans les prochaines sections, j'explore la trajectoire historique qui a permis à Patterson de défendre, en 1982 cette conception bien précise de l'homologie synonyme de synapomorphie.

En phylogénétique, pour que la similarité implique l'homologie, il faut minimalement qu'elle soit tributaire d'une ascendance commune (Hall 2007; Kapli, Yang et Telford 2020; Morrison, Morgan et Kelchner 2015). Cette notion traduit ainsi la posture philosophique selon laquelle toute catégorie biologique légitime doit se conformer à la théorie de l'évolution par voie

de sélection naturelle et ainsi témoigner de l'origine évolutive commune des organismes ou des traits qu'on regroupe, et ce, à tous les niveaux d'organisation (Griffiths 2007; Hull 1970, 2001; Millikan 1984, 2002; Neander 1995, 2002). Le critère selon lequel la similarité, pour être dite homologique, doit être porteuse d'un signal phylogénétique, c'est-à-dire qu'elle doit relever de l'ascendance commune des organismes portant les traits impliqués, est tout de même imprécis. Cette imprécision explique pourquoi les systématiciens et systématficiennes parlent d'un « problème » de l'homologie (de Beer 1971; Inkpen et Doolittle 2016; Panchen 1994).

En effet, considérant que tous les êtres vivants ont un ancêtre commun (qui pourrait être une communauté d'organismes), toute similarité est potentiellement significative sur le plan phylogénétique. Le « problème » de l'homologie est donc celui d'établir des critères pour distinguer les similarités qui sont des homologies de celles qui ne le sont pas ou peu (j'évite, à partir d'ici, de complexifier la discussion en parlant d'homologies partielles, mais il est important de garder en tête que toutes les similarités phylogénétiquement importantes ne le sont pas nécessairement au même degré; Haggerty et al. 2014; Kendig 2016). Le problème est bien sûr toujours celui de la coordination entre hypothèses ou énoncés théoriques et phénomènes : ce qui est théoriquement accepté comme étant la cause de l'homologie (ascendance commune des espèces mises en relation) est inobservable de manière directe.

Le concept d'homologie incarne ainsi la coordination des données disponibles pour analyse biologique avec la théorie qui encadre le traitement de ces données afin de générer des systèmes classificatoires hiérarchiques. La chaîne inférentielle concernée peut être résumée en trois étapes. D'abord, les descriptions des organismes et des espèces offrent des données quantitatives et qualitatives au sujet de traits morphologiques. Ces descriptions sont utilisées pour formuler des hypothèses d'homologie. Par exemple, il est possible de défendre, sur la base de leur similarité, que les traits A et B, portés respectivement par les espèces *a* et *b*, sont des homologues. De telles hypothèses sont formulées sur la base de différents critères, les trois les plus communs étant le principe de correspondance topologique (placement relatif des traits par rapport au reste de l'organisme), le principe de connexion (existence d'états de caractère intermédiaires reliant les deux traits observés) et la structure même des traits observés (Griffiths 2007; Remane 1952; Rieppel 2015; Williams 2004). En plus de ces trois critères, l'inscription de

traits au sein de mécanismes développementaux similaires (critère ontogénique) est parfois mobilisée (Brigandt 2003; Panchen 1992; Wagner 1989). Dans les raisonnements impliquant ces critères d'analyse, il est implicite que l'homologie est irréductible à la similarité brute ou à la découverte de l'identique. En reprenant un phrasé classique, il s'agit bien d'identifier des structures qui méritent le même nom malgré leurs différences (Panchen 1994). Cette première étape de l'analyse permet d'identifier ce qui a été appelé des homologies premières (de Pinna 1991) ou des homologies topographiques (Jardine 1969), c'est-à-dire des relations de similarité détenant potentiellement un signal phylogénétique.

Ensuite, une fois que les hypothèses sont ainsi formulées, elles peuvent être testées, ce qui permet d'identifier des homologies secondaires (de Pinna 1991) ou phylogénétiques (Jardine 1969). La logique impliquée est intimement liée au paradigme darwinien. En effet, si l'homologie a bien été identifiée, alors elle devrait se conformer au patron généalogique des espèces, dénotant une relation de similarité au sein d'un groupe monophylétique. Pour corroborer la présence d'une telle correspondance au patron généalogique, il convient de tester la cohérence entre différentes homologies. Dit de manière plus contemporaine, si plusieurs hypothèses d'homologies utilisées convergent sur un seul et même patron phylogénétique, celui-ci peut être considéré comme robuste; on considère alors que les différentes hypothèses d'homologie se renforcent mutuellement, qu'elles sont congruentes (Patterson 1982; Wiley et Lieberman 2011; Williams 2004). Cette étape est typiquement appelée « test de congruence » et mobilise la congruence non seulement entre les données étudiées, mais aussi celle relative aux systèmes taxonomiques déjà existants (Panchen 1992; Patterson 1982; Rieppel 2015; Williams 2004).

La croyance selon laquelle l'identification d'homologies topographiques ou premières est athéorique est bien sûr inadéquate (la présence des quatre critères d'analyse de la similarité fournis précédemment témoigne d'une charge théorique des observations), mais il pourrait être argumenté qu'elle ne mobilise pas la théorie de l'évolution (Brower 2000; Hull 1970, 2001; Sokal et Sneath 1963; Williams 2004). Au contraire, le test de congruence est incontestablement évolutionniste, puisque les théoriciens et théoriciennes de la cladistique traditionnelle considèrent que l'évolution entraîne la formation d'un patron, un ordre au sein duquel chaque groupe taxonomique n'occupe qu'une seule place :

Avoiding further examples, my general point is that of order versus disorder. If several homologies characterize the same group, as hair, mammary glands, ear ossicles, left systemic arch, etc. characterize Mammalia, we have overwhelming evidence of order (Riedl [1978]), and currently assume that order, or pattern, to be the product of phylogeny (history). (Patterson 1982, 42)

Ainsi, toutes les homologies tendent à raconter la même histoire puisqu'elles sont issues de l'évolution qui sous-tend la phylogénie. Si, au contraire, les espèces avaient été créées indépendamment les unes des autres, un tel ordre pourrait être absent; des relations de similarité pluridirectionnelles pourraient lier les organismes sans qu'il ne soit théoriquement justifié d'en prioriser certaines et d'en délaissier d'autres. La théorie de l'évolution par voie de sélection naturelle fournit ainsi le cadre métaphysique qui permet aux phylogénéticiens et aux phylogénéticiennes de tester la validité d'une homologie postulée.

Finalement, une fois corroborées par congruence, les homologies sont mobilisées pour reconstruire des systèmes phylogénétiques, eux-mêmes utilisés pour former des catégories (les taxons supérieurs à celui d'espèce) et des systèmes taxonomiques. En ce sens, l'homologie peut être décrite comme un principe qui incarne la hiérarchie des systèmes classificatoires : « Homologous features and conditions form a hierarchy that corresponds to the taxonomic hierarchy of groups » (Bock 1963, 268). Cette façon de systématiser est opérée par l'identification d'homologies de plus en plus générales. Par exemple, la main humaine est homologue à celle du chimpanzé, mais aussi, dans une perspective plus générale, à la patte du chat (et plus spécifiquement à celle de mon chat et collègue d'écriture, Lou). Dans cette perspective bien précise, l'homologie est une relation de similarité qui circonscrit des groupes monophylétiques, mais qui s'exprime à différents niveaux d'analyse, étant plus ou moins inclusive selon l'endroit où l'on se situe dans ce jeu de perspectives.

Ce court résumé du processus inférentiel démontre la centralité de la notion d'homologie pour la biologie comparative en général et la phylogénétique tout particulièrement. Il s'agit de l'axe sur lequel pivote toute inférence, du concept qui unit similarité et généalogie, qui fait le pont entre observations plus directes et reconstructions phylogénétiques : « In other words, homology is the relation that systematists and comparative anatomists use in generating hypotheses of relationship. » (Patterson 1982, 29). Plus récemment, Fitch réitère cet engagement envers la

teneur évolutive de la notion d'homologie qui permet de la distinguer, au moins conceptuellement, de l'analogie (Fitch n'étant pas philosophe, vous excuserez son utilisation inexacte du terme « abstraction ») : « Homology is here an abstraction in that it is a relationship, common ancestry, the nature which we find important to know about » (Fitch 2000, 228).

Il convient maintenant de bien identifier la circularité qui, d'un point de vue anhistorique, semble émerger des raisonnements entourant l'homologie (la circularité typique du problème de la coordination). Bien que l'homologie ait par ailleurs souvent été définie en référant à la notion de similarité, les théoriciens et théoriciennes de la cladistique ont, dans la deuxième moitié du 20^e siècle, cherché à éviter cette voie qui rend explicite la circularité en question :

Perhaps [...] a reason for deleting similarity from the evolutionary definition of homology is to escape the charge of circularity in definition – similarity as used to infer the common ancestry, which is used to validate the similarity as homology. (Patterson 1982, 24)

Voici un exemple de définition qui témoigne de telles tentatives d'éviter la circularité en question : « Features (or conditions of a feature) in two organisms are homologous if they stem phylogenetically from the same feature (or the same condition of the feature) in the immediate common ancestor of these organisms. » (Bock 1977, 881). L'idée sous-jacente est qu'une référence à l'ascendance commune permet d'éviter la circularité au sein de la définition. Cette stratégie est comparable à l'application d'une couche de peinture sur une portion rouillée de carrosserie. Cela ne règle aucunement le problème ainsi masqué, comme en témoigne d'ailleurs l'occurrence du terme « same » dans la définition de l'homologie précédente.

En effet, que la similarité soit présente dans la définition de l'homologie ou non, toute mise en œuvre du concept l'implique nécessairement. Le concept d'homologie en phylogénétique est précisément ce qui fait le lien entre étude morphologique (ou génétique) et inférences phylogénétiques. Autrement dit, l'homologie ne peut remplir son rôle inférentiel crucial sans mobiliser la similarité : celle-ci est théoriquement pertinente, c'est-à-dire qu'elle devient homologique, si et seulement si elle dénote un héritage généalogique. Williams met l'accent sur ce point en insistant sur la qualité phénétique des données mobilisées par la phylogénétique (les matrices de données auxquelles il réfère dans ce qui suit sont formées de

données issues d'analyses de similarité; pour rappel, la phénétique est un courant de la biologie systématique qui rejette la phylogénie comme critère de classification) : « Little has been said about [...] the recognition that current binary data matrices are essentially phenetic, no more than mere lists of things that are similar. » (Williams 2004, 13-14). Tant que cette réalité pratique reste d'actualité, l'omission définitionnelle de la similarité demeure un geste conceptuellement futile; le rapport de réciprocity entre similarité et phylogénie, circulaire d'un point de vue anhistorique et par lequel s'incarne le problème de la coordination, est toujours présent. Inkpen et Doolittle abondent en ce sens :

Mindell and Meyer (2001), in a review with goals similar to ours, suggest that similarity criteria be abandoned: they "can be helpful in diagnosing, but not defining homology." But we do need such criteria, because without them we are unable (in both conceptual and empirical senses) to identify or delimit traits (at any level) whose homology is to be assessed. That is, we cannot begin to ask if two features are homologs unless we have some way to single them out as features to be compared, that is some measure of similarity (minimally, for instance, that both meet some definition of "gene"). (Inkpen et Doolittle 2016, 191)

Dans ce qui suit, je cherche à démontrer, par une historiographie, que cette relation entre similarité et phylogénie, que j'ai décrite jusqu'à présent de manière anhistorique et circulaire, est cependant plus adéquatement représentée par une hélice, un cercle étiré sur un axe temporel. Cette relation inférentielle étant à la base de toute pratique phylogénétique, je défends l'idée selon laquelle la structure épistémique de la discipline semble circulaire seulement lorsque sa dimension diachronique est niée, ce qui obscurcit une partie du phénomène d'intérêt pour l'épistémologie, soit la structure épistémique en question. Pour ce faire, je décompose schématiquement l'histoire de l'homologie en trois stades : l'époque prédarwinienne, la réinterprétation darwinienne du concept et sa sophistication telle qu'elle a été illustrée par les travaux de Hennig, un troisième temps qui se clôt sur l'établissement de la synonymie liant homologie et synapomorphie.

1.6.2 L'homologie prédarwinienne

Selon l'idée reçue, Richard Owen (1804-1892) est considéré comme le premier systématicien à avoir clairement identifié la relation d'homologie comme celle qui doit guider la classification. Plus précisément, il distingue les traits homologues des traits analogues. Selon sa

définition qui précède la publication de la théorie darwinienne de l'évolution, deux organes sont homologues s'ils représentent la « même » structure sous différentes formes ou fonctions (Owen 1843). Les deux organes sont les « mêmes » en ce sens qu'ils devraient porter le même nom (*namesakes*), malgré les différences qui les caractérisent. Typiquement, deux organes méritent de porter le même nom si et seulement s'ils sont des équivalents matériels d'un organe postulé au sein de l'archétype. L'archétype est une représentation idéalisée de la structure des organismes, le produit d'une abstraction auquel les formes concrètes se conforment plus ou moins bien (il s'agit ici de l'acception qu'Owen attribue tardivement à la notion d'archétype et à laquelle les historiens et historiennes réfèrent le plus souvent; voir Panchen 1994 pour le détail de l'évolution de la notion au cours de la carrière d'Owen). En contraste, la notion d'analogie est réservée aux structures qui ont la même fonction, qu'elles soient des homologies ou non, c'est-à-dire qu'elles soient liées à l'archétype de la même manière ou non.

Trois types d'homologies sont distinguées par Owen : sérielle, spéciale et générale. L'homologie sérielle réfère à deux organes au sein d'un seul et même organisme, évoquant du même coup le concept de paralogie de la biologie moléculaire d'aujourd'hui. Les deux organes « en série » sont associés au même organe au sein de l'archétype. L'homologie spéciale est quant à elle la relation entre organes homologues présents au sein de deux organismes. En ce sens, les ailes des chauves-souris et les bras qui me permettent d'écrire ces quelques mots sont homologues. L'homologie générale traduit la relation entre un trait au sein d'un organisme et l'organe correspondant au sein de l'archétype.

Chez Owen, le pouvoir explicatif de l'homologie dépend de l'existence d'un archétype à partir duquel des lois de structure peuvent être identifiées (Panchen 1992, 1994; Rieppel 2015; van Valen 1982). C'est bien *parce qu'il y a un archétype*, duquel les formes particulières sont des approximations, que certaines structures en viennent à porter le même nom. De surcroît, deux organes liés par une homologie spéciale se ressemblent de manière pertinente d'un point de vue classificatoire parce qu'ils entretiennent chacun une homologie générale avec la même structure archétypique.

Prise de manière anhistorique, la structure argumentative d'Owen est indéniablement circulaire : l'archétype permet de distinguer analogies et homologies, alors que son existence ne peut qu'être déduite des relations de similarité entre organismes puisqu'il est inobservable. Cependant, le contexte sociohistorique témoigne du fait qu'Owen inscrit ses recherches au sein d'une longue tradition de systématiciens et de systématficiennes, une tradition qui donne du poids à chacun des éléments de sa structure justificative. En effet, Macleay (1792-1865) théorisait dès 1821 l'idée selon laquelle il faut distinguer les similarités qui offrent un regard sur l'ordre naturel (qu'il appelait affinités) de celles qui sont moins cruciales au travail des naturalistes (analogies) dans le cadre de travaux en taxonomie (Macleay 1821). Strickland (1811-1853) considérait d'ailleurs en 1846 que Macleay, et non Owen, était à créditer pour la mise en place du concept d'homologie (Panchen 1994; Strickland 1846).

Au-delà de ce litige historiographique, l'idée d'une telle distinction entre simple similarité et similarité qui justifie que deux organes portent le même nom s'inscrit dans une histoire qui est, de toute façon, bien plus longue que celle qui s'amorcerait avec Macleay ou Owen, même dans le cadre restreint de la classification des êtres biologiques. Par exemple, il y a plus de deux mille ans, Aristote classait les dauphins avec les mammifères, de sorte que l'idée d'homologie est au moins implicite dans sa logique classificatoire (Rieppel 2015). Panchen (1994) fait aussi remarquer que la notion d'homologie, dans le sens qu'Owen lui attribue, est implicite dans les travaux de Goethe (1749-1832), d'Oken (1779-1851) et de Carus (1789-1849). Cela dit, avant le duo Macleay-Owen, l'illustration la plus nette du concept d'homologie est certainement celle présente dans les travaux de Belon (1517-1564). Dans une planche anatomique bien connue, Belon illustre de manière on ne peut plus claire le fait que des squelettes distincts, comme celui d'un humain et celui d'un oiseau, se conforment approximativement à un plan d'organisation unique. Par la juxtaposition des deux squelettes disposés similairement, il soulignait des relations d'homologie (sans utiliser cette expression) en suggérant que des structures ayant des fonctions bien distinctes, comme les os des ailes et ceux des bras, puissent porter la même étiquette. Il s'agissait bien d'un premier élan d'anatomie comparée au sein de la culture intellectuelle occidentale, tout aussi crucial au développement du concept d'homologie que la cristallisation de sa définition par Owen presque 300 ans plus tard (Kendig 2016; Panchen 1994).

La notion d'archétype, intimement liée à celle d'homologie, précède elle aussi les travaux d'Owen (voir Rupke 1993 pour une étude détaillée de l'histoire préowenienne du concept), bien que les références philosophiques de celui-ci à des auteurs comme Platon soient plutôt éparses et manquent de clarté (Panchen 1994). Nonobstant cette confusion, l'idée d'une correspondance des formes matérielles et particulières à un plan abstrait et idéalisé (dans une perspective platonicienne) ou même l'interprétation des spécificités qui les caractérisent comme étant accidentelles (dans une perspective aristotélicienne) s'impose très certainement par son ancienneté. On peut notamment associer la notion d'archétype à l'unité du plan organisationnel chez Étienne Geoffroy de Saint-Hilaire (1772-1844), ainsi qu'à l'influence que des auteurs comme Cuvier (1769-1832) et von Baer (1792-1876) ont eue sur Owen (Rupke 1993).

La distinction nette entre analogie et homologie a su rendre concrète la différence entre simple similarité et similarité qu'il est théoriquement justifié de mobiliser en systématique, offrant par le fait même un concept autour duquel pouvait se sophistiquer la biologie comparative. Il convient donc de parler d'Owen comme de l'un des fondateurs de la systématique moderne, mais il serait erroné de concevoir ses travaux comme un point de départ de l'histoire de la systématique ou même de l'histoire du concept d'homologie. Comme tout juste exposé, ces histoires remontent au moins jusqu'aux balbutiements de la tradition scientifique occidentale, trouvant leur terme provisoire dans l'absence de sources remontant plus loin (Morange 2017).

1.6.3 L'homologie et la mise en place de la phylogénétique

Il est souvent réitéré que les travaux de Darwin, malgré l'onde de choc évolutionniste qu'ils ont générée, ont laissé la systématique largement intacte. Cette idée est à la fois exacte et imprécise. Les systèmes taxonomiques sont demeurés largement inchangés à la suite de la publication de *L'Origine des espèces* (1859). Cependant, le cadre qui les soutient, qui leur donne un sens, qui guide les recherches subséquentes et qui ancre les méthodes d'analyse a quant à lui été complètement bouleversé. L'homologie permet d'illustrer la teneur de ce nouvel ancrage théorique, ainsi que son effet sur la pratique, et éventuellement, sur le va-et-vient historique entre pratique et théorie qui caractérise l'histoire de la phylogénétique.

Pour Owen, l'homologie est à comprendre par l'entremise du rôle explicatif de la notion d'archétype : deux homologues (une homologie spéciale les relie) témoignent d'une similarité importante puisqu'ils entretiennent la même relation à l'archétype (l'homologie générale relie chaque homologue à l'archétype). Pour Darwin, c'est la notion d'ancêtre commun qui vient, par l'entremise de la perspective évolutionniste, jouer ce rôle de clé de voûte conceptuelle. Deux organes sont considérés comme étant « le même » (c'est-à-dire comme étant des homologues) malgré les différences (de forme, de fonction, de structure, etc.) s'ils sont dérivés du même organe présent au sein d'un ancêtre commun des organismes qui les portent :

[S]uppose that the ancient progenitor, the archetype as it may be called, of all mammals, had its limbs constructed on the existing general pattern, for whatever purpose they served, we can at once perceive the plain signification of the homologous construction of the limbs throughout the whole class. (Darwin 1859, cité dans Kendig 2016, 113)

L'unité du plan organisationnel est dorénavant expliquée par l'unité d'ascendance. Nous savons aujourd'hui que cette réinterprétation de la notion d'archétype fait partie d'un discours scientifique qui a eu et a encore aujourd'hui des impacts massifs en sciences comme dans le reste de nos cultures intellectuelles. Cela dit, la pratique taxonomique, dont les systèmes étaient pourtant au cœur de l'analyse de Darwin, n'en demeure pas moins, au moment de la publication de *L'Origine*, privée de nouvelles méthodes qui auraient été définies par la pensée évolutionniste. En effet, bien que la théorie darwinienne ait exhorté les taxonomistes à classer les organismes selon leurs liens généalogiques traduits par l'homologie, ceux-ci demeurent inobservables. Il faut donc les inférer, et pour ce faire, les mêmes critères que ceux utilisés par les biologistes de l'ère prédarwinienne sont mobilisés : l'ontogénie, la topologie des organes et la connectivité permettent d'identifier les liens de similarité porteurs d'un signal phylogénétique. Les systèmes taxonomiques passés sont donc conservés, bien qu'ils aient été produits par une vision fixiste des espèces. Plutôt que d'être rejetés, ces systèmes servent de base justificative pour l'argument darwinien qui leur confère un nouveau sens : la théorie de l'évolution par voie de sélection naturelle est valide, argumente Darwin, notamment puisqu'elle explique mieux le patron identifié par les taxonomistes, soit des regroupements hiérarchiques au sein desquels on retrouve des groupes d'espèces se ressemblant plus les unes aux autres qu'elles ne ressemblent aux espèces issues d'autres groupes (voir chapitres 2 et 4 de *L'Origine des espèces*).

La notion d'homologie¹⁰ s'inscrit à merveille dans cette logique de reconstruction de systèmes hiérarchiques, mais il sera nécessaire de la raffiner avant qu'elle ne permette d'identifier des groupes monophylétiques, tâche que les cladistes hennigiens et hennigiennes lui attribuent (homologie taxique). Les travaux de l'inventeur du terme *phylogénie*, Haeckel (1834-1919), témoignent bien des complications relatives à la conversion évolutionniste des notions de systématique. Au sein des nombreux arbres phylogénétiques qu'il a offerts, Haeckel inclut régulièrement des groupes paraphylétiques, c'est-à-dire des groupes excluant certains des descendants d'une espèce ancestrale donnée (par exemple, en excluant les oiseaux du groupe qu'ils devraient former avec les reptiles; Willmann 2003). C'est donc dire que certaines espèces qui, sur une base généalogique, auraient pu être classées avec les autres descendantes d'une espèce ancestrale sont exclues du groupe. Cette pratique est d'autant plus surprenante que Haeckel mobilisait la notion d'homologie pour reconstruire l'ordre d'apparition des groupes monophylétiques (terme dont il est aussi l'inventeur) : « Evidently one must distinguish what Haeckel meant by monophyly from what he required of taxa. » (Farris 1990, 82; voir aussi Panchen 1992, Rieppel 2016, Willmann 2003).

Ce choix semble être le résultat d'un respect marqué pour les systèmes taxonomiques déjà disponibles que Haeckel vient réinterpréter en termes généalogiques, les représentant sous la forme d'arbres. La réinterprétation de l'homologie en termes d'un passé évolutif commun aux homologues (l'unité de type expliquée par l'unité d'ascendance) est toujours insuffisante, chez Haeckel, pour balayer les résultats taxonomiques passés, même si elle mènera éventuellement à la maturation de la phylogénétique. La possibilité de considérer des groupes plus inclusifs (ou moins inclusifs, selon la perspective adoptée) que les groupes monophylétiques est ainsi toujours ouverte, malgré la présence d'un cadre théorique qui invite justement à se concentrer exclusivement sur ces groupes monophylétiques.

¹⁰ Le terme d'homologie n'est pas utilisé par Darwin, contrairement à la notion d'homologue. Afin de simplifier la discussion, j'utilise ici la notion d'homologie pour référer à la relation entre les homologues, comme si Darwin avait utilisé ce même raccourci sémantique.

Cette situation fait écho aux débats, beaucoup plus récents, concernant la place des archées au sein de l'Arbre du vivant. En effet, bien que la transition vers la phylogénétique moléculaire ait rapidement fourni des données suggérant que les protistes (ou procaryotes), un groupe traditionnellement composé de ce qu'on appelle aujourd'hui les archées et les bactéries, forment un groupe paraphylétique (les eucaryotes, proches parents des archées, en étant exclus) qu'il convient donc de diviser en deux groupes monophylétiques (les archées et les bactéries), cette idée aujourd'hui bien ancrée dans la discipline a fait face à une résistance marquée à ses débuts, une résistance assise sur le respect des systèmes taxonomiques passés (Woese, Kandler et Wheelis 1990; Woese 1994).

Cette transition de la systématique vers la systématique phylogénétique témoigne d'un autre impact significatif imputable à la théorie darwinienne, soit la cristallisation d'une posture métaphysique relative aux systèmes classificatoires. Pour des auteurs comme Linné ou, plus tardivement, Bentham (1800-1884) et Hooker (1817-1911), les systèmes de classification hiérarchiques sont artificiels (les espèces elles-mêmes étant tout de même considérées comme réelles). Dans cette optique, l'objectif des systèmes taxonomiques est d'épauler la mémoire dans sa fonction permettant aux scientifiques de retrouver l'information au sujet d'objets d'étude donnés (Stevens 1997; Tort 1989). Il est donc tout à fait plausible, dans cet ordre d'idées, que les systèmes hiérarchiques issus de la classification soient distincts de l'ordre naturel. Chez Linné, cette possibilité est actualisée alors que le naturaliste suédois nous a offert des taxonomies hiérarchiques et les outils de nomenclature qui y sont associés tout en considérant que l'ordre naturel était mieux traduit par l'idée d'une échelle de la nature, la fameuse *scala naturae* (Lovejoy 1936).

Au contraire, des auteurs tels qu'Adanson (1727-1806), Antoine-Laurent de Jussieu (1748-1836) Cuvier ou Agassiz (1807-1873) considéraient que les systèmes taxonomiques devaient se conformer à l'ordre naturel (Tort 1989). C'est dans cette veine qu'il faut comprendre la métaphysique darwinienne : l'ordre naturel est donné par les relations généalogiques entre organismes. Les organismes les plus proches phylogénétiquement forment des variétés. Une variété dont les membres ont du succès au sein de la lutte pour l'existence en vient à être caractérisée par des traits plus marqués ainsi qu'à croître en nombre de sorte qu'elle puisse

éventuellement être considérée comme une espèce. Une espèce peut, de la même manière, se différencier toujours plus des autres espèces (tout particulièrement de son espèce sœur), et par ailleurs, voir en son sein naître de nouvelles variétés qui pourront éventuellement former des espèces, puis des genres, des ordres, des familles, etc. Ce processus d'évolution qui est ancré dans la sélection naturelle ainsi que dans le principe de divergence des lignées qui l'accompagne implique qu'il n'y a pas de différence métaphysique entre espèces et taxons supérieurs : tout rang représente ultimement une lignée monophylétique. C'est donc dire que les systèmes hiérarchiques classificatoires devraient éventuellement calquer l'ordre naturel; celui-ci est un système phylogénétique.

Cette calcification d'intuitions réalistes au sujet des systèmes classificatoires (réalistes en ce sens qu'elles suggèrent que les systèmes traduisent l'ordre naturel) par l'entremise de la théorie darwinienne de l'évolution a été d'une importance capitale pour le développement de la phylogénétique. L'homologie a ainsi été appelée à devenir plus qu'un lien de similarité théoriquement pertinent entre traits morphologiques; elle est devenue le principal outil de *découverte* de l'histoire évolutive : « Following Darwin, homology was seen not just as evidence of evolution, but also as the way to build phylogenetic relationships. » (Inkpen et Doolittle 2016, 186). Elle représente à la fois un idéal, soit celui d'un lien de similarité porteur de signal phylogénétique univoque, et un concept pouvant être opérationnalisé pour la reconstruction phylogénétique. La fin du 19^e et une bonne partie du 20^e siècles permettent ainsi d'observer comment cet outil en vient à tendre vers cet idéal et les complications qui émergent sans surprise de cette sophistication du concept.

1.6.4 La monophylie comme cible de la phylogénétique et critère définitionnel de l'homologie

La notion d'homologie (taxique) est centrale pour la cladistique, sous ses formes hennigiennes ou subséquentes. Dans ce contexte, elle est associée à la relation entre deux groupes taxonomiques reliés par un ancêtre commun *immédiat*. Des homologues forment ainsi un groupe monophylétique restreint, et ces groupes sont les blocs avec lesquels on reconstruit l'histoire des espèces, des gènes, des organismes, etc. Or, cette réduction de l'homologie à ce qui

est aussi appelé la synapomorphie ne suit qu'indirectement l'idée darwinienne selon laquelle « l'unité de type est causée par l'unité d'ascendance ». Il a donc fallu, dans l'espoir d'extraire l'histoire évolutive à partir de traits observables, peaufiner les concepts en jeu et leur utilisation.

Un premier pas important dans cette direction a été d'identifier nettement le manque à gagner, sur le plan de la précision, qui accompagne la notion théoriquement chargée d'homologie. Au tournant du 20^e siècle, Lankester (1847-1929) et Mivart (1827-1900) considéraient par exemple que la notion d'homologie était trop large et polysémique pour servir les inférences phylogénétiques. Lankester suggère de conserver le terme d'homologie, mais pour référer de manière abstraite à toute similarité *potentiellement* phylogénétiquement porteuse. Après que des homologies ont été testées sur la base de la congruence, comme décrit plus tôt (Jardine 1969; de Pinna 1991; Rieppel 2015), afin d'identifier si la similarité en question témoigne bien d'une proximité généalogique reliant les homologues, deux types de relations plus précises peuvent être distinguées. Si le test se révèle concluant, l'homologie, selon Lankester, peut se révéler être une homogénie (*homogeny*). L'homogénie est, en pratique, l'équivalent de l'homologie phylogénétique (lien de similarité entre deux traits qui témoignent de la présence d'un ancêtre commun). Si, au contraire, le test de congruence échoue, alors l'homologie se révèle plutôt être une homoplasie, un terme introduit par Lankester et dont la signification est demeurée similaire au travers du temps (Panchen 1994; Williams 2004). Une homoplasie est une similarité morphologique incohérente avec d'autres données phylogénétiques, de sorte qu'on doive en attribuer la cause à autre chose qu'une ascendance commune aux organismes qui portent les traits en question. Finalement, le terme d'homoplasie a marqué la discipline, mais le glissement vers l'homogénie comme remplaçante de l'homologie, au contraire, a avorté. C'est plutôt la consécration de l'acception phylogénétique de l'homologie qui marque cette époque, alors que le terme est de plus en plus utilisé dans le sens que Lankester attribuait à l'homogénie (Williams 2004).

Il faut par la suite attendre la deuxième moitié du 20^e siècle pour que des développements conceptuels marqués viennent enrichir le concept d'homologie. D'abord les travaux d'Hennig, puis leur absorption par l'ensemble de la communauté scientifique occidentale (au-delà de l'Allemagne), engendrent l'introduction de nouveaux termes pour penser le rapport entre

similarité et généalogie. Il s'agit de bien qualifier dans quelle mesure la similarité peut être utilisée pour inférer la phylogénie : « Therefore it is not the extent of resemblance or difference between various organisms that is of significance for research into phylogenetic relationship, but the connection of the agreeing or divergent characters with earlier conditions. » (Hennig 1965, 102). Au nombre des concepts phares de la cladistique hennigienne, on retrouve les suivants : d'une part, la monophylie, la paraphylie et la polyphylie pour délimiter des groupes et, d'autre part, les relations entre les traits qui accompagnent ces délimitations, soit respectivement la synapomorphie, la simplésiomorphie et la convergence (ou, plus largement, l'analogie).

La monophylie hennigienne, la synapomorphie et la paraphylie ont été décrites précédemment, mais je me permets ici un court rappel. Un groupe est dit monophylétique s'il inclut tous et uniquement les descendants d'une espèce ancestrale; les synapomorphies permettent de les cerner. Un groupe est dit paraphylétique s'il inclut seulement des descendants d'une espèce ancestrale sans tous les inclure; les relations simplésiologiques entre caractères le délimitent. Les simplésiologies sont des états de caractère hérités d'un ancêtre commun mais ayant disparu ou s'étant transformés drastiquement chez certains descendants de cet ancêtre. Ainsi, les simplésiologies ne lient pas tous les descendants de l'ancêtre en question, mais seulement ceux qui ont conservé ledit état de caractère. Par exemple, on peut considérer les reptiles comme étant un exemple paradigmatique de groupe paraphylétique, et le fait qu'ils soient ectothermes (à sang froid) dénote la présence d'un trait simplésiologique puisque les oiseaux, qui complètent le groupe monophylétique, sont endothermiques (à sang chaud).

Finalement, les groupes polyphylétiques incluent des organismes ou espèces issus de différentes lignées; la relation de similarité reliant ses membres est appelée homoplasie, le terme introduit par Lankester faisant preuve d'une pérennité enviable. Les homoplasies sont à éviter lorsqu'il est question de reconstruire le passé phylogénétique des groupes à l'étude sur la base de relations de similarité puisqu'elles sont le produit d'une convergence et non d'une divergence. En d'autres termes, la cladistique considère que ces relations de similarité sont dénudées de tout signal phylogénétique (Hennig 1965). Conséquemment, la cladistique hennigienne, qui stipule une relation univoque entre classification et phylogénie (ce qui, historiquement, n'est pas le cas

de l'ensemble de la systématique évolutionniste) rejette aussi la valeur classificatoire des groupements circonscrits par les homoplasies.

Ce court survol des notions phares de la cladistique hennigienne illustre bien que le concept évolutionniste d'homologie, c'est-à-dire le concept reliant en son cœur la similarité et la phylogénie en référant à l'ascendance commune, est décomposé par Hennig. Avant ses travaux, ce concept était trop vague pour servir de base inférentielle et classificatoire, comme mentionné plus tôt. Hennig en extrait donc des concepts qui cernent des relations plus précises, permettant des reconstructions phylogénétiques plus fines. La synapomorphie est la clé de la cladistique hennigienne; il s'agit bien d'une relation homologique qui permet d'identifier, contrairement aux simplésiomorphies et aux apomorphies, des groupes monophylétiques.

Ce genre de distinction permet de répondre au problème observé dans les travaux de Haeckel qui incluait au sein de ses systèmes classificatoires des regroupements d'espèces paraphylétiques. Or, pour Hennig et la majorité des praticiens et des praticiennes d'aujourd'hui, un tel geste classificatoire est douteux. La théorie de l'évolution par voie de sélection naturelle les exhorte à classifier sur la base du patron arborescent qui traduit la trajectoire divergente des espèces, une divergence opérée par rapport à l'ancêtre commun. Ainsi, il s'agit de reconstruire un schème hiérarchique qui rend compte de l'ordre naturel formé de lignées indépendantes qui marquent leur divergence de manière toujours plus forte (Hennig 1965). En traitant les reptiles comme un groupe taxonomique, par exemple, on suggère implicitement que ceux-ci représentent une branche de l'arbre du vivant, ce qui est faux.

On peut reformuler la logique qui précède comme suit : puisque les inférences phylogénétiques partent d'hypothèses formulées notamment sur la base de systèmes taxonomiques préalables, il faut pouvoir reconnaître les groupes qui ont été identifiés à tort comme étant phylogénétiquement adéquats. Certains de ces groupes lacunaires sont paraphylétiques et ils doivent être corrigés afin que la classification n'inclue que des groupes monophylétiques. C'est du moins ainsi que l'on doit lire le projet de classification phylogénétique, comme informé par la cladistique hennigienne, projet ayant vu la relation d'homologie se morceler et, du même coup, se préciser.

Ce fractionnement ne marque cependant pas le terme de l'évolution du concept d'homologie (ni le terme de son morcellement). Sans compter le fait que l'homologie est au cœur de la phylogénétique moléculaire (qui est elle-même au cœur de l'analyse de cette thèse à partir du chapitre 5), elle a aussi su maintenir sa place au sein de la phylogénétique morphologique s'étant élaborée à la suite des travaux de Hennig. Patterson, en 1982, offre une analyse détaillée du concept d'homologie tel qu'il est alors utilisé. Il identifie quatre types de définitions du concept : classique (celle d'Owen, par exemple), évolutionniste (l'homologie, en ce sens, postule une transformation d'un trait vers un autre, ne se limitant donc pas à identifier des relations entre traits), phénétique (non phylogénétique) et cladistique (taxique). Pour le dire simplement, l'homologie cladiste est l'équivalent de la synapomorphie et il s'agit du concept que Patterson (1982) et d'autres cladistes retiennent (Brigandt 2003; Griffiths 2007; Rieppel 2015; Williams 2004).

1.6.5 Cohérentisme historique et hélice phylogénétique

L'histoire de l'homologie s'écrit encore aujourd'hui, prenant la forme d'une symphonie polysémique. Malgré l'hétérogénéité de cette notion et des usages, la plupart des praticiens et des praticiennes s'entendent pour dire que des homologues sont des traits biologiques comparables (similaires ou identiques) et tributaires d'une histoire évolutive commune (Hall 2007; Kapli, Yang et Telford 2020; Wiley et Lieberman 2011). À ce niveau de généralité, l'homologie n'est évidemment plus synonyme de synapomorphie. En revanche, peu importe le niveau d'abstraction adopté, le rôle logique du concept au sein des pratiques demeure le même : traduire le lien entre similarité et phylogénie. C'est du moins ce qu'une analyse anhistorique de chacun des moments de l'histoire de l'homologie permet de déceler.

Le survol de l'histoire de l'homologie qui couvre la période allant de la première utilisation du concept jusqu'aux années 1980 permet aussi d'identifier d'autres éléments cruciaux de la structure épistémique de la phylogénétique. Premièrement, et c'est le point principal que je cherche ici à illustrer, cela met en lumière la relation d'influence mutuelle entretenue par les aspects théorique et concret des pratiques classificatoires. Les systèmes taxonomiques du 18^e siècle et de la première moitié du 19^e siècle forment la condition même de possibilité de la

formulation du concept d'homologie par Owen et de la théorie de l'archétype qui encadre ce concept et qui lui donne un sens. En effet, sans ces systèmes taxonomiques préalables qui privilégiaient certaines relations de similarité, Owen n'aurait eu ni la motivation ni la légitimité épistémique pour structurer ainsi ses idées, qui ont par la suite contribué à raffiner la pratique.

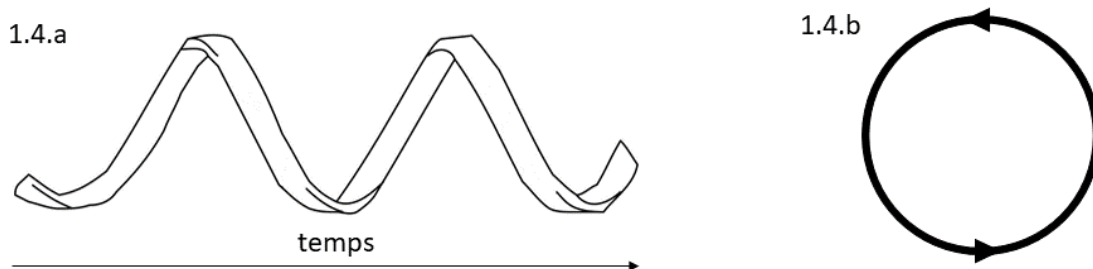
Ce type de va-et-vient se reproduit : les pratiques taxonomiques et la notion d'homologie ont contribué à faire émerger le paradigme darwinien qui se mettait tranquillement en place alors même qu'Owen théorisait la distinction entre homologie et analogie : « Thus, well before Darwin dared announce his theory in public, some of the most powerful supporting evidence for that theory had already been established as biological orthodoxy. » (Griffiths 2007, 647).

Finalement, plus de cent ans après *L'Origine*, la notion d'homologie s'est précisée progressivement, étant éventuellement définie, sous l'influence de la cladistique hennigienne, comme la relation entre deux traits hérités d'un ancêtre commun et immédiat des espèces qui les portent (Patterson 1982). Cet aboutissement théorique (et provisoire) vient structurer les inférences phylogénétiques des cladistes de l'époque et, surtout, les systèmes classificatoires qui en découlent, ceux-ci étant dès lors basés (exclusivement) sur la monophylie.

Ce va-et-vient entre théorie et pratiques taxonomiques qui permet la coordination entre théorie et empirie évite toutefois la circularité; il détient en effet une dimension diachronique et les transformations successives qui caractérisent la structure épistémique de la phylogénétique témoignent du fait que la circularité est le fruit d'une analyse décontextualisée. Il s'agit ici de souligner que le problème de la coordination entre théories et phénomènes est bien présent tout au cours de cette histoire. Rien ne peut changer la situation épistémique basique de la phylogénétique, celle-ci devant nécessairement jongler avec des liens de similarité observables et des relations généalogiques inobservables. C'est cependant plutôt une hélice, comme un brin d'ADN orphelin, qui permet d'illustrer le plus adéquatement la structure épistémique impliquée, et non le cercle, comme cela a été longuement défendu en philosophie de la mesure (voir chapitre 2). Chaque itération du patron circulaire (présent dans chaque réflexion phylogénétique isolée de son contexte sociohistorique) est différente du reste (un fait invisibilisé par une description circulaire des pratiques scientifiques). Elle s'appuie non pas sur elle-même (comme la circularité

le suggère), mais sur ce qui la précède historiquement, sur la légitimité du complexe discursif duquel elle fait partie. La Figure 1.4 illustre ce modèle hélicoïdal.

Figure 1.4 – Deux représentations de la structure justificative des pratiques scientifiques



Cette figure illustre deux modèles permettant de représenter la structure justificative des pratiques scientifiques. En 1.4.a, le modèle prend la forme d'une hélice. Le va-et-vient entre théorie et procédures d'observation y engendre un mouvement circulaire. Ce mouvement circulaire est toutefois étiré sur un axe temporel, ce qui transforme le cercle en hélice et permet de représenter explicitement le changement. Chaque itération épistémique (isolée pragmatiquement) est ainsi représentée par un mouvement circulaire complet. Le tout témoigne d'une continuité au sein de la structure épistémique : les plus récentes itérations s'appuient sur les plus anciennes. En 1.4.b, on observe le même mouvement circulaire qui représente le va-et-vient entre théorie et procédures d'observation. Cette seconde représentation fait abstraction de la dimension temporelle de la structure épistémique, ce qui rend difficile la conceptualisation du dynamisme inhérent aux pratiques scientifiques.

Deuxièmement, tracer l'histoire de l'homologie permet d'évaluer la stabilité de la structure épistémique étudiée. Le concept d'homologie et le cadre évolutionniste ont eu un lourd impact sur les pratiques classificatoires, les rendant d'abord évolutionnistes, puis axées sur la monophylie. Or, l'entièreté de cette toile épistémique est encadrée par la capacité à formuler des hypothèses d'homologie (puis de synapomorphie si la synonymie est abandonnée). Il faut alors insister sur le fait notable selon lequel les liens de similarité biologiquement importants, tant que nous restons dans le domaine morphologique, sont identifiés sur la base de critères stables, et ce, depuis le tournant du 19^e siècle, donc depuis l'époque prédarwinienne. Les critères préconisés sont la topologie, la connectivité, la structure et l'ontogénie. En référant à Agassiz, un biologiste notoirement anti-darwinien, Patterson résume la situation comme suit : « in the words of a great nineteenth century paleontologists (Agassiz 1874, 92) "It is worth our while to ask if there is any such process as evolution in nature. Unquestionably, yes. But all that is actually known of this process we owe to the great embryologists." Agassiz then names von Baer. » (Patterson 1982,

65). Tout ce qu'on connaît du processus évolutif, nous disent Agassiz et Patterson, est issu de l'analyse anatomique (fortement influencée par l'embryologie au 19^e siècle). L'histoire de la phylogénétique est ainsi celle qui donne un sens historique à des observations contemporaines, un accomplissement impressionnant d'un point de vue épistémologique, qui rend l'inaccessible observable, du moins selon les phylogénéticiens et phylogénéticiennes. Dans un texte plus récent portant sur la polysémie du terme « homologie », Griffiths (2007) a similairement défendu l'idée selon laquelle la diversité des concepts de l'homologie renvoie en fait à diverses *explications* d'un phénomène unique, soit la similarité établie sur la base des critères de Remane (identité topologique, principe de connexion et spécificité structurelle des homologues).

Or, la formulation d'hypothèses d'homologie n'est que la première étape de la reconstruction phylogénétique et de la classification des espèces. Une fois ces hypothèses formulées, il faut les tester par l'entremise du test de congruence décrit plus tôt pour savoir s'il s'agit bien d'homologies (*contra* Griffiths). Une fois leur validité confirmée (ou infirmée) par la résolution des incongruences, les homologies (*sensu* Patterson et l'approche taxique) servent de base pour l'inférence phylogénétique et les systèmes taxonomiques. Les différents modèles alors utilisés voient eux aussi leur validité s'ancrer dans la théorie darwinienne de l'évolution. En cladistique, par exemple, le modèle arborescent est nécessaire pour restreindre le champ des possibles, celui-ci étant, comme illustré précédemment, intimement lié à la théorie darwinienne. Cette affirmation tient la route malgré le fait que les patrons hiérarchiques associés au modèle arborescent précèdent de beaucoup les travaux de Darwin (on peut tracer l'histoire du modèle arborescent au moins jusqu'à l'arbre de Porphyre et donc jusqu'à la méthode analytique néo-platonicienne; Crombie 1994; Panchen 1992) puisque la seule justification pour l'utilisation de ces modèles qui est de nos jours valide et recevable dans le champ phylogénétique est leur adéquation avec l'ordre naturel tel qu'il est suggéré par la théorie de l'évolution par voie de sélection naturelle (dans sa forme traditionnelle).

Deux conclusions doivent être tirées quant à la stabilité du système épistémique étudié. D'abord, la suggestion de Patterson et Griffiths selon laquelle les connaissances sur le processus évolutif pourraient être réduites aux critères permettant d'identifier l'homologie est inadéquate puisque les inférences ne font que débiter avec ces identifications; la structure épistémique de

la phylogénétique est traversée par des croyances évolutionnistes qui nous informent sur les processus historiques à différents niveaux. Cette structure est donc moins stable que ces deux auteurs ne le suggèrent.

Ensuite, lorsque certaines croyances constitutives d'une toile épistémique sont remises en question, cette toile doit être révisée (et non détruite). Le chapitre 7 analyse la mise en pratique d'une telle remise en question, soit la remise en question localisée de la validité du modèle arborescent en phylogénétique. L'analyse épistémologique permet alors de saisir toute l'importance de ces changements sur l'ensemble du complexe discursif. Au chapitre 6, en contraste, j'analyse un débat phylogénétique au sein duquel les itérations successives sont plus similaires. Nonobstant cette différence de degré, les deux cas d'étude gagnent à être considérés en tenant compte de l'historicité de la structure épistémique puisque celle-ci permet de souligner et d'appréhender la complexité des relations qu'entretiennent les différentes croyances composant l'aspect théorique d'un champ scientifique et d'identifier leur rapport toujours changeant aux phénomènes.

Autrement dit, pour bien analyser les relations entre les croyances constitutives d'un système épistémique, il faut connaître leur histoire, même si l'analyse visée se veut logique et anhistorique. Des liens qui semblent nécessaires, comme celui entre similarité et phylogénie ou entre évolution darwinienne et modèle arborescent, peuvent dès lors offrir un nouveau visage. Adopter une vision historique de l'objet d'étude de l'épistémologie permet ainsi d'opérer une analyse critique *en plus* de considérer une facette du phénomène trop souvent ignorée (en philosophie des sciences anglo-saxonne ou analytique), soit son historicité.

1.7 Conclusion

Les deux objectifs principaux de ce chapitre étaient premièrement d'illustrer la présence du problème de la coordination au sein des pratiques phylogénétiques, et deuxièmement d'illustrer l'historicité de la structure épistémique de ce champ d'étude, caractérisée par un va-et-vient entre pratiques concrètes (analyse de similarité, polarisation des traits biologiques, constructions de systèmes phylogénétiques et classificatoires, etc.) et théoriques (définitions de concepts, élaboration de nouvelles théories, transformation de théories existantes,

développement de modèles, etc.). Le premier objectif a été accompli par une analyse de la cladistique hennigienne qui souligne que toute tentative d'inférer le passé généalogique d'organismes sur la base de la similarité requiert l'acceptation apriorique de cette relation inférentielle. Celle-ci est toutefois impossible à corroborer indépendamment des observations qu'elle rend possibles. Cette logique inférentielle peut être associée à l'ensemble de la phylogénétique qui s'articule irrémédiablement sur le lien similarité/phylogénie.

Le deuxième objectif a été atteint par l'étude historique du concept d'homologie, dont l'histoire est celle d'un concept forgé à partir de systèmes taxonomiques existants, transformé par le darwinisme et devenu, une fois raffiné, apte à guider les pratiques classificatoires. Puisque l'homologie traduit la relation inférentielle similarité/phylogénie, son histoire est métonymiquement celle de la phylogénétique que je considère, par extension, comme étant mieux représentée par un modèle hélicoïdal que par un modèle circulaire. Le va-et-vient entre théorie et empirie a bien une dimension historique, du moins dans le cas de la phylogénétique.

Ces deux constats guident l'analyse épistémologique des pratiques phylogénétiques contemporaines que je réalise aux chapitres 5, 6 et 7 de cette thèse. Il est donc crucial d'en comprendre la portée philosophique. Pour le dire simplement, le problème de la coordination implique l'impossibilité de considérer l'expérience comme étant le fondement absolu de la connaissance. Au contraire, l'expérience, telle qu'elle est mobilisée en sciences, est un phénomène chargé de théorie, d'éléments théoriques devant être acceptés *a priori* et sans lesquels aucune donnée ne peut être acquise. Si l'empirisme classique et fondationnaliste est indubitablement remis en question par cette situation épistémique, il faut éviter de confondre ce constat avec celui d'une menace pour la connaissance scientifique. Bien au contraire, les philosophes de la mesure ont réconcilié le fait indéniable qu'est l'existence du problème de la coordination avec l'empirisme de manière à guider l'analyse épistémologique. Dans le chapitre suivant, j'explore ces travaux philosophiques.

Chapitre 2 – Vers une philosophie de la classification inspirée du problème de la coordination et de la philosophie de la mesure

To learn which questions are unanswerable, and not to answer them:
this skill is most needful in times of stress and darkness.

- Ursula K. Le Guin, dans *The Left Hand of Darkness*

2.1 Introduction

Le chapitre précédent ayant démontré l'importance d'appréhender les dynamiques de justification scientifique en tenant compte de leur dimension diachronique et du problème de la coordination, je présente, dans ce chapitre-ci, un courant théorique que j'appelle le cohérentisme historique et qui répond à cette injonction. Plus précisément, je démontre que les approches de Chang (2004), van Fraassen (2008) et Tal (2012, 2016), parmi d'autres approches développées en philosophie de la mesure (Frigerio, Giordani et Mari 2010; Mari 2000, 2005; Morrison 2009; Morrison et Morgan 1999) et que j'englobe aussi au sein du cohérentisme historique, fournissent des outils indispensables à la bonne compréhension et à la description précise des dynamiques scientifiques. Aux chapitres 5, 6 et 7, ces outils sont mobilisés pour étudier les pratiques phylogénétiques contemporaines.

La défense du cohérentisme historique, que je présente dans les pages qui suivent, s'articule autour des implications épistémiques et épistémologiques du problème de la coordination. À l'instar de Chang, van Fraassen et Tal, je fais de ce casse-tête mon point de départ. Il convient donc de m'intéresser aux différentes approches qui, au cours des cent cinquante dernières années, ont traité du problème de la coordination. Je cherche à démontrer que le cohérentisme historique est le courant de pensée le plus apte à rendre compte de l'impact du problème de la coordination sur les pratiques scientifiques *tout en accordant un rôle important à l'expérience* au sein de la toile épistémique des sciences.

Le chapitre est divisé en deux parties. Je survole dans un premier temps les approches de philosophie de la mesure du 20^e siècle qui entrent en dialogue avec le problème de la coordination, soit le conventionnalisme, le procéduralisme et le constructivisme. Il s'agit bien d'un survol centré sur le problème de la coordination et ses ramifications; il n'est aucunement question d'une revue du corpus de la philosophie de la mesure dans son ensemble. Ainsi, les questionnements relatifs à l'incertitude au sein des processus de mesure, la typologie des échelles utilisées, la teneur métaphysique des quantités physiques et théoriques, l'individuation des quantités, l'histoire des sciences métrologiques, par exemple, ne sont mentionnés que s'ils sont nécessaires à la compréhension du propos entourant le problème de la coordination. Dans le même ordre d'idées, les exemples de biologie et de phylogénétique utilisés dans ce chapitre servent à l'exposition des thèses de nature philosophique. Une exploration plus détaillée de certains de ces exemples est nécessaire à la compréhension des pratiques phylogénétiques; ce travail est accompli plus loin dans la thèse (chapitres 5, 6 et 7), une fois que les éléments théoriques de philosophie sont mis en place.

Après avoir identifié les faiblesses des approches traditionnelles au problème de la coordination, je montre, dans un deuxième temps, comment certains auteurs contemporains (Chang 2004; Tal 2016; van Fraassen 2008), que j'associe au cohérentisme historique, tissent une épistémologie empiriste qui surpasse ces lacunes. Pour conclure cette deuxième partie, je présente une critique des approches cohérentistes historiques, formulée récemment dans une optique explicitement réaliste, et j'offre une défense de ce courant de pensée ayant grandement influencé mon travail.

2.2 Les réponses philosophiques au problème de la coordination (20^e siècle)

2.2.1 Le conventionnalisme

Les premiers auteurs ayant souligné le va-et-vient lié à la coordination des théories avec l'empirie sont Mach et Poincaré, considérés comme les fondateurs de l'approche conventionnaliste (Poincaré 1898; Mach 1896; Tal 2012; van Fraassen 2008). Pour Mach, la

mesure est une opération comparative qui implique une part de convention, notamment lorsqu'un standard, c'est-à-dire un point de comparaison partagé par les membres d'une communauté, est utilisé comme unité de mesure. Pour Poincaré, la science est structurée de manière hiérarchique; cette hiérarchie est fondée sur des conventions, qui sont en fait des définitions théoriques de base. Il s'agit de principes non arbitraires, mais impossibles à vérifier directement de manière empirique, comme le principe d'inertie, nous dit Poincaré.¹¹ Nonobstant l'insaisissabilité empirique de ces conventions, elles sont mobilisées pour guider la recherche, notamment en structurant les hypothèses, qui peuvent quant à elles être confirmées ou infirmées empiriquement.

C'est donc dire que l'idée de convention renvoie à quelque chose de distinct chez chacun de ces auteurs. Pour Mach, ce sont les standards de mesure qui sont établis par convention. Pour Poincaré, les conventions sont des définitions théoriques qui sont par la suite mobilisées pour rendre l'observation possible. Cependant, les deux auteurs se rejoignent sur un point important : l'acceptation apriorique de notions symboliques (définitions ou standards) est nécessaire pour toute opération métrologique (Mach 1896; Poincaré 1898; Tal 2015; van Fraassen 2008). Le conventionnalisme offre ainsi une réponse au problème de la coordination selon laquelle la connaissance empirique trouve son fondement dans les conventions définitionnelles *qui lui sont extérieures*; il s'agit d'une approche fondationnaliste (Chang 2007). Cette approche, en acceptant l'inévitabilité du va-et-vient accompagnant la coordination (voir chapitre 1, tout particulièrement les sections 1.2 et 1.3), rejette de fait l'idée selon laquelle les croyances scientifiques pourraient être justifiées empiriquement par l'observation pure (dénudée de théorie), mais conserve la possibilité de fonder la connaissance de manière absolue. Les écrits de Reichenbach (1927), dans les mains de qui s'est cristallisé l'approche conventionnaliste, en témoignent clairement.

Face au problème de la coordination, Reichenbach accepte de fait, comme ses prédécesseurs, que toute vérification empirique d'une théorie scientifique est circulaire (comme

¹¹ Le principe d'inertie est un élément fondamental de la physique classique qui stipule, chez Newton par exemple, que les corps physiques maintiennent leur mouvement en l'absence de forces externes. Ils tendent donc à résister aux forces qui influencent leur mouvement (DiSalle 2020).

démontré au chapitre précédent avec la cladistique hennigienne). Cela dit, il considère que cette circularité est strictement limitée aux *modalités* des pratiques de mesure; leurs conditions de possibilité, au contraire, peuvent être conceptualisées sans qu'une situation aussi fâcheuse survienne. Pour Reichenbach, la possibilité même de la coordination d'un phénomène et d'énoncés théoriques repose sur des principes de coordination aprioriques non soumis à la vérification empirique, comme chez Poincaré. Il s'agirait en fait de *principes définitionnels* qui établissent *a priori* la régularité du phénomène à mesurer et, de ce fait, la possibilité de les mesurer (Reichenbach 1927; van Fraassen 2008).

Reichenbach s'intéresse notamment à la mesure du temps, une situation qui permet de bien illustrer la notion de principe définitionnel. Comme chez Mach, la dimension conventionnelle de la mesure ne relève que du choix d'horloge. On peut, par exemple, souhaiter utiliser la Terre comme horloge et sa rotation sur elle-même comme unité standard de mesure. Lorsque nous utilisons la Terre pour mesurer le temps, nous considérons que ses périodes (une rotation complète de la Terre sur elle-même) sont régulières. Chaque période se déroulant sur un intervalle de temps égal au suivant, nous pouvons quantifier la durée d'un phénomène selon le nombre de périodes de l'horloge qui sont réalisées pendant que le phénomène suit son cours. C'est donc dire que la bonne mesure du temps repose sur l'adoption d'une horloge effectivement régulière. Une question s'impose dès lors : comment s'assurer que l'horloge choisie est fiable?

Cette tâche est plus difficile qu'on pourrait le croire puisqu'il est impossible de comparer les périodes successives d'une horloge, celles-ci se déroulant de fait à des moments différents. Il est impossible de les juxtaposer pour les comparer et donc de s'assurer que, par exemple, deux journées ont bien la même durée. Il en découle potentiellement une certaine gêne épistémique à utiliser la période terrestre ou un segment de celle-ci pour mesurer le temps.

De plus, lorsque des incohérences entre des résultats de mesure surviennent (immanquablement), la source est difficile à identifier : cela peut être dû au phénomène mesuré lui-même, qui fluctue; cela peut être dû à des problèmes dans la procédure de mesure (l'horloge et son utilisation); les incohérences métrologiques peuvent aussi être dues à des fluctuations dans le cours du temps lui-même. La situation est exacerbée par l'utilisation d'une deuxième horloge

visant à évaluer la régularité de la première. En effet, l'établissement de la régularité de cette deuxième horloge souffrant des mêmes difficultés, les diagnostics qu'elle peut générer sont aussi peu déterminés que les diagnostics portant sur les résultats incohérents de la première horloge. Il y a donc sous-détermination empirique des procédures métrologiques liées au temps.

Pour Reichenbach ainsi que les praticiens et les praticiennes, ces difficultés sont contournables par l'acceptation de croyances aprioriques. Le nœud épistémique tout juste décrit se dissout de lui-même si la régularité d'une horloge est posée comme étant une vérité sémantique, une définition première à partir de laquelle il devient possible d'ériger diverses pratiques empiriques et discours gnoséologiques. Ainsi, Reichenbach met l'accent sur l'impossibilité de démontrer que des principes comme celui de la régularité du temps sont démontrables empiriquement, mais il double cette affirmation d'une séparation stricte entre le royaume de la connaissance et celui de la sémantique qui lui permet de faire reposer le premier sur le second (la « détermination impossible » dont il parle au début de la citation suivante est celle de la régularité du temps) :

Why is this determination impossible? Do not the laws of physics, for instance those of the motion of a pendulum, compel us to believe in the equality of the periods? It is true that the laws as described in textbooks suggest this belief; but if we ask ourselves where these laws come from, we shall find that they are obtained through observations of clocks calibrated according to the principle of the equality of their periods. The proof is therefore circular. [...] A solution is obtained only when we apply our previous results about spatial congruence and introduce the concept of a *coordinative definition* into the measure of time. The equality of successive time intervals is not a matter of knowledge but a matter of *definition*. (Reichenbach 1927, 116)

Tout se passe ici comme si la circularité omniprésente et pourtant bien marquée par Reichenbach ne serait qu'un épiphénomène négligeable par rapport au principe reliant observation et théorie par le langage. La sous-détermination des procédures métrologiques serait ainsi le fait de l'établissement d'une unité et de procédures de mesure, mais la possibilité même de la mesure serait quant à elle surdéterminée par la possibilité d'établir des principes définitionnels (il ne pourrait pas en être autrement).

Concrètement, le conventionnalisme suggère ainsi de nier l'importance épistémologique du problème de la coordination en mettant l'accent sur un geste définitionnel qui serait

logiquement premier, ayant préséance épistémique sur la mise en place de procédures concrètes. *Ex nihilo*, une première définition conventionnelle, celle établissant de fait la régularité du temps ou d'une horloge par exemple, réalise les conditions de possibilité de la coordination et de la mesure. Cela dit, aucun des auteurs traditionnellement associés au conventionnalisme (Mach, Poincaré, Reichenbach, Carnap, Ellis) ne fournit de reconstruction historique qui nous permettrait d'identifier factuellement de tels moments définitionnels, que ce soit pour la mesure de la température, du temps ou d'une autre quantité. En conséquence, il faut concevoir le conventionnalisme comme étant une réponse atemporelle au problème de la coordination : *en principe*, il pourrait y avoir un tel moment définitionnel, un principe sémantiquement premier et sous-jacent qui viendrait établir les bases nécessaires à la connaissance (Chang 2007; van Fraassen 2008). Pourtant, dans les faits, comme le reconnaît Reichenbach lui-même dans un passage déjà cité, les définitions ont bel et bien une histoire : « but if we ask ourselves where these laws come from, we shall find that they are obtained through observations of clocks calibrated according to the principle of the equality of their periods. » (Reichenbach 1927, 116).

Ainsi, le conventionnalisme, en fournissant une réponse ahistorique au problème de la coordination, néglige complètement le fait que les outils et procédures reliés aux pratiques de mesure sont des objets historiques. Ils sont dotés d'une histoire qui se révèle être structurée par un rapport mutuel et pérenne entre observations et définitions, entre expérience et théorie :

At the same time [Reichenbach] means to be still actively inquiring into the empirical conditions under which such definitions can play the requisite role. What he did not do is change the ahistorical setting of his problem: the coordination is still apparently to be conceived of as possible in the absence of any previous such coordination. But is that possible at all? We have to ask more or less the same question again as before: how can such coordinative definitions be meaningfully introduced except in a historical context where there are some prior coordinations already in place? I submit that they cannot. (van Fraassen 2008, 121)

L'isolement du geste sémantique premier par les auteurs conventionnalistes doit dès lors être considéré comme étant une abstraction qui occulte ce processus historique dont la substance semble pourtant importante pour la compréhension des dynamiques scientifiques (Chang 2004; Morrison et Morgan 1999; Morrison 2009; Schaffer 1992; Shapin et Schaffer 1985; Tal 2016; van Fraassen 2008).

Ce point a été illustré au chapitre précédent par un regard porté sur l'histoire de la notion d'homologie et son rapport à la phylogénétique. En effet, j'ai d'abord démontré que dans une perspective anhistorique, la phylogénétique hennigienne semble avoir une structure épistémique circulaire : elle mobilise des principes imbriqués dans la théorie de l'évolution par voie de sélection naturelle (par exemple, la croyance selon laquelle les liens de parenté entre espèces sont corrélés à la similarité morphologique) afin de contribuer à la confirmation empirique de cette même théorie (par la reconstruction d'arbres phylogénétiques et l'identification de groupes monophylétiques). Dans une perspective conventionnaliste, ces éléments théoriques doivent être considérés comme étant extérieurs à la connaissance, comme étant des principes sur lesquels la connaissance reposerait. On pourrait ainsi, à la manière de Hennig, émettre des principes aprioriques et les mobiliser pour inférer la phylogénie des organismes et des espèces, sans se soucier de la circularité qui en résulte.

Concrètement, ce raisonnement tend à circonscrire l'arène des débats : puisque les principes aprioriques sont des vérités sémantiques, elles sont hors de portée des débats scientifiques courants; elles sont plutôt à concevoir comme étant des contraintes fixes à ces échanges. Évidemment, des remises en question de ces principes sont possibles, même en adoptant une posture conventionnaliste, mais cela implique de fait une transition scientifique significative et difficile à engendrer. Comme je le détaille plus loin, le conventionnalisme est obsolète en philosophie de la mesure. Je défends qu'une analyse philosophique de la phylogénétique (et éventuellement d'autres pratiques classificatoires) devrait emboîter le pas. Au sein de la phylogénétique elle-même, ce type de raisonnement est toutefois d'actualité et considéré comme tout à fait viable, tel que je le démontre aux chapitres 6 et 7.

Les auteurs conventionnalistes traitent certaines croyances scientifiques comme étant des fondations nécessaires pour éviter les problèmes liés à la coordination, une perspective qui acquiert son sens lorsque les systèmes épistémiques sont considérés de manière anhistorique. Or, j'ai démontré, au chapitre précédent, que l'étude anhistorique de la phylogénétique, qui sert ici d'exemple, occulte un élément important de sa structure justificative, soit, justement, son dynamisme historique. *Chacun des éléments d'une structure épistémique qu'on parvient à isoler par analyse détient une historicité propre, inscrite dans celle plus large du système auquel elle*

appartient et qui informe son rôle dans ce système. En négligeant ce fait, on perd nécessairement des informations cruciales. L'approche que je préconise, soit le cohérentisme historique, embrasse au contraire cette historicité et analyse les croyances scientifiques à la lumière de leurs relations justificatives, certes, mais aussi de leurs relations historiques qui offrent un regard plus nuancé (et précis) sur les soi-disant fondations épistémiques. En poursuivant mon analyse des réponses philosophiques au problème de la coordination, je démontre dans le reste du chapitre la validité de cette approche soucieuse de l'histoire.

Avant de ce faire, il convient cependant de compléter la description des approches conventionnalistes en explorant ses ramifications dans la deuxième moitié du 20^e siècle. Malgré son rejet qui est généralement accepté en philosophie de la mesure, le conventionnalisme, après avoir été réactualisé par Carnap (1966) et Ellis (1966), a acquis une certaine pérennité par l'entremise de la théorie représentationnelle de la mesure (*Representational Measurement Theory*; Diez 1997, 1997; Krantz et al. 1971; Suppes et al. 1989; Luce et al. 1990). Celle-ci peut être considérée comme une réhabilitation des approches fondationnalistes en philosophie de la mesure. Cette théorie considère l'attribution d'une valeur pour décrire un phénomène comme étant l'établissement d'une relation d'homomorphie entre une variable théorique et un phénomène ou un aspect d'un phénomène (Krantz et al. 1971; Mari 2000). Cette relation est à comprendre au sens de l'approche sémantique des théories scientifiques, c'est-à-dire que la quantité établie est une *représentation* d'un aspect du phénomène. Cette quantité est valide si et seulement si elle entretient des relations aux autres éléments symboliques de la théorie qui seraient comparables aux relations que l'aspect du phénomène qui a été mesuré entretient avec d'autres aspects matériels de ce phénomène. Par exemple, la quantité théorique *température* entretient, au sein de la thermodynamique, la même relation à la quantité théorique *volume* que la température d'un phénomène entretient avec son volume.

La théorie représentationnelle (contrairement à l'approche de van Fraassen dont les travaux s'inscrivent aussi en cohérence avec l'approche sémantique) contourne les enjeux relatifs à la coordination de la théorie et des phénomènes puisqu'elle accepte de fait la possibilité de la relation homomorphique sans appréhender les modalités de la coordination entre un phénomène et sa représentation scientifique, une coordination nécessaire au constat

d'homomorphie. Les contributions de cette théorie à la philosophie de la mesure consistent principalement à appréhender les types de relations de vérité qui existent entre les résultats de pratiques métrologiques et les phénomènes mesurés ainsi qu'à établir une typologie des échelles de mesure (Tal 2012).

La théorie représentationnelle de la mesure défend, du moins implicitement, une posture fondationnaliste comparable à celle de Reichenbach, Mach et Poincaré. En effet, la validité de la connaissance repose encore une fois sur quelque chose qui lui est extérieur, soit l'existence de phénomènes observables avec lesquels les théories peuvent entretenir une relation d'homomorphie. Cette relation d'homomorphie implique une dualité qui évoque l'histoire de l'épistémologie occidentale : d'un côté, la connaissance, de l'autre, la matérialité, les deux étant liées par l'observation. Autrement dit, un problème épistémologique ancestral issu du dualisme est reformulé, sans toutefois que sa charge métaphysique ne soit explicitée. Dans cette logique, les deux domaines sont indépendants, de sorte qu'il serait tout à fait possible d'établir la validité des connaissances en s'appuyant confortablement sur les phénomènes (ou ce qui les sous-tend). Comme chez Reichenbach, la circularité du problème de la coordination est ainsi reléguée aux modalités de la mesure, sans pour autant éclabousser l'ensemble de l'édifice épistémique de l'empirisme : « [*The representational measurement theory*] assumes the empirical properties as the a priori elements on which the scale construction, and then the measurement itself, can be founded. » (Mari 2005, 262, italiques ajoutés). La question de savoir comment on peut effectivement construire les échelles et pratiques métrologiques sur la base de propriétés empiriques auxquelles ces échelles et pratiques donnent accès est cruciale pour toute épistémologie des sciences, mais ignorée par la théorie représentationnelle de la mesure.

C'est donc dire que, tout comme dans le cas du conventionnalisme, l'aboutissement fondationnaliste est atteint seulement parce que le va-et-vient diachronique entre théorie et observation est écarté. Si celui-ci, au contraire, était considéré comme le cœur du phénomène étudié, soit la structure épistémique d'un ensemble de pratiques scientifique, alors toute forme de fondationnalisme empirique semblerait vouée à l'échec. En effet, qu'elle mobilise des principes définitionnels ou une relation d'isomorphie entre phénomènes et énoncés théoriques,

la posture fondationnaliste sous-estime la contribution des dynamiques sociohistoriques à la construction de connaissances scientifiques.

Malgré cet angle mort saillant, les philosophies associées au conventionnalisme ainsi qu'à la théorie représentationnelle de la mesure ont su attirer l'attention des philosophes et métrologues sur un point déterminant pour la philosophie de la mesure et pour toute philosophie empiriste : même si l'explication de la normativité scientifique est irréductible à la présence d'une fondation atemporelle et absolue à la connaissance, les pratiques scientifiques sont jugées, du moins par les praticiens et praticiennes, comme étant plus ou moins précises et comme ayant plus ou moins de succès. En d'autres termes, la normativité (le fait que certains modèles, pratiques, théories, procédures, articles, livres, standards, etc. sont considérés comme étant supérieurs à d'autres) est un fait des dynamiques scientifiques qu'il serait maladroit de négliger. Toutefois, le problème de la coordination, contrairement à ce que le conventionnalisme suggère, est bel et bien l'élément structurel principal des phénomènes de justifications empiriques qui sous-tendent cette normativité, de sorte qu'il faille réconcilier les réponses aux deux problèmes (celui de la normativité et celui de la coordination) et non en développer une au profit de l'autre.

2.2.2 Opérationnalisme

Alors que la réponse conventionnaliste se concrétise sous la plume de Reichenbach, Bridgman (1882-1961), physicien lauréat du prix Nobel de 1946, développe une approche philosophique de la mesure au sein de laquelle les procédures matérielles occupent une place prépondérante. Bridgman publie, en 1927, *The Logic of Modern Physics*. Dans cet ouvrage, l'auteur s'interroge notamment sur le lien entre les concepts théoriques, comme celui de température, et les procédures qui permettent de les observer concrètement (procédures permettant la coordination entre phénomène et théorie). Pour lui, ce lien est conceptuellement simple : un concept théorique se définit exclusivement par l'entremise des procédures qui permettent de l'observer. Si un concept est privé de procédures d'observation, il est dénué de sens. Dans le même ordre d'idées, chaque procédure implique son propre concept, de sorte que deux manières différentes de mesurer la température sonderaient en fait les phénomènes en fonction de deux quantités théoriques distinctes.

Selon cette approche, le problème de la coordination implose. En effet, celui-ci repose sur la dichotomie entre théorie et phénomènes doublée de l'impossibilité de mesurer avec satisfaction un phénomène sans l'acceptation apriorique d'éléments théoriques, eux-mêmes fondés sur l'expérience. Or, pour Bridgman, les procédures sont développées pour répondre à des problèmes pratiques et en viennent ainsi à orienter le travail théorique, mais ce rapport est unidirectionnel. La circularité est évitée puisque la théorie n'a selon lui aucune légitimité pour informer la mise en place de procédures métrologiques et circonscrire les concepts quantitatifs. Si le procéduralisme a reçu sa part de soutien au moment de sa formulation par Bridgman, il rencontre toutefois nombre de difficultés qui découlent, pour le dire simplement, de sa naïveté. Les deux problèmes principaux du procéduralisme sont les suivants : d'abord, il implique une multiplication illégitime des concepts, puis il ignore complètement le problème de la coordination.

La multiplication des concepts impliqués par le procéduralisme est problématique de deux façons : elle est normativement indésirable et inapte d'un point de vue descriptif. Chang démontre par exemple que la croyance apriorique selon laquelle le concept de température s'applique au-delà des champs d'application stabilisés de la thermométrie (au-delà des phénomènes ayant des températures modérées) a été, du moins au 19^e siècle, un moteur efficace pour le développement de nouveaux instruments et de procédures permettant de mesurer les températures extrêmes. Ces développements ont notamment permis de *confirmer* la croyance préalable selon laquelle les températures extrêmes et modérées pouvaient être subsumées par un concept unique. L'opérationnalisme, dans sa version stricte, rendrait impossible ce genre de mouvement exploratoire, confinant chaque concept à son champ d'application *déjà* acquis (Chang 2004).

Cette situation témoigne du fait que l'unité d'un concept auquel des pratiques d'observation distinctes sont associées peut être épistémiquement bénéfique. Il serait alors nuisible de guider les dynamiques scientifiques à l'aune du procéduralisme de Bridgman. De surcroît, ignorer le fait que praticiens et praticiennes, bien qu'ils et elles mesurent la température de différentes façons, croient en l'unité du concept serait, d'un point de vue descriptif, erroné. La

discussion entourant la notion d'homologie moléculaire, au chapitre 4 de cette thèse, démontre que le même argument s'applique à la phylogénétique.

Bridgman avait toutefois envisagé les conséquences potentiellement néfastes de la multiplication des concepts. Ainsi, il fournit un deuxième critère d'individuation des quantités théoriques (le premier étant leur association univoque à des procédures). Si deux ensembles de procédures offrent des résultats généralement cohérents (ou qui ne témoignent pas d'incohérence), alors il est adéquat de considérer que ces méthodes identifient la même quantité. Cette solution est cependant naïve puisque la cohérence entre des méthodes, en plus de toujours être approximative, est le fruit de labeurs qui cherchent à établir leur cohérence sur la base de leur rapport à un concept unique. Il est donc problématique de considérer la cohérence entre les méthodes comme étant un point de départ pour l'individuation des quantités théoriques : « this criterion is naive, because different operations that are thought to measure the same quantity rarely agree with each other before being deliberately corrected for systematic errors » (Tal 2012, 69).

La convergence des méthodes, nous dit Tal, est l'aboutissement de pratiques et non leur point de départ. Il est en effet impossible de déterminer empiriquement si des procédures distinctes convergent ou non, sinon par l'entremise de manipulations théoriquement chargées. Par exemple, pour savoir si une colonne de mercure et une colonne d'alcool offrent des résultats cohérents, il faut détenir des fonctions permettant de lier les deux procédures distinctes à la quantité théorique qu'est la température. Le lien entre température et procédures distinctes sert de guide qui permet de mettre à jour une éventuelle cohérence. En d'autres termes, l'identification de la cohérence entre deux méthodes d'observation dépend de la croyance apriorique de l'unité du concept théorique impliqué. Cette identification échoue donc à servir de critère d'individuation des quantités théoriques de manière à répondre au problème de la multiplication des concepts au cœur du procéduralisme.

Force est d'admettre qu'une telle erreur est facilement évitée par toute épistémologie qui place à l'avant-plan le problème de la coordination et la charge théorique des observations, ce qui confirme du même coup le deuxième problème du procéduralisme. La faille de l'approche de

Bridgman est en ce sens patente : elle stipule qu'en se fiant aux procédures seules, la définition des quantités théoriques est donnée de fait et liée tout naturellement à des phénomènes concrets. Cette approche néglige le fait omniprésent selon lequel les procédures reposent sur des développements théoriques préalables et que même la délimitation des phénomènes à observer est issue d'un rapport à la théorie (Chang 2004; Frigerio, Giordani et Mari 2010; Mari 2005; Tal 2017; van Fraassen 2008).

Pour ces raisons (parmi d'autres; voir van Fraassen 2008), je considère l'opérationnalisme très peu adapté à la description fine des pratiques scientifiques, et plus spécifiquement des pratiques phylogénétiques. Les concepts théoriques gagnent à être définis autrement que par une référence univoque à un ensemble de procédures. D'un point de vue normatif, l'injonction de Bridgman à se méfier de la polysémie des concepts peut toutefois guider d'importantes recherches de clarté. S'il paraît en effet trop contraignant de systématiquement nier la synonymie d'occurrences d'un terme sur la base de procédures différentes auxquelles elles sont associées (pratique épistémologique beaucoup trop contraignante pour l'analyse de la phylogénétique, comme je le détaille au chapitre 4), il pourrait être tout à fait approprié de viser la rencontre entre la pratique et la théorie afin de s'assurer que l'utilisation d'un terme comme celui d'homologie ne soit pas galvaudée par l'accumulation de couches sémantiques. Pour ce faire, il convient toutefois de prendre en compte la complexité de la toile épistémique d'un ensemble de pratiques plutôt que d'en abstraire artificiellement une simplicité comme celle que Bridgman aurait bien voulu rendre utile.

2.2.3 Le constructivisme et le cohérentisme

Le cohérentisme est considéré comme la principale option de rechange au fondationnalisme pour rendre compte des dynamiques de justification scientifique et pour les inscrire à la base d'une épistémologie empiriste (Chang 2007). Là où les approches fondationnalistes cherchent à établir une base absolue à la connaissance, habituellement cherchée du côté de l'expérience (les fondationnalismes idéalistes ayant été préalablement écartés de la philosophie occidentale au tournant du 20^e siècle), le cohérentisme affirme que ce qui permet d'expliquer la validité des connaissances est à trouver dans la cohérence du système

épistémique qui les porte. Selon Chang, les diverses attaques (Foucault 1966; Longino 1990; Lloyd 1995; Quine 1961, 1971; Sellars 1956; Wittgenstein 1969) qu'ont subies les tentatives variées de développer une épistémologie empiriste considérant l'expérience isolée comme fondation de la connaissance nous incitent fortement à adopter une perspective cohérentiste (Chang 2007). En effet, la deuxième moitié du 20^e siècle a vu se généraliser, au sein de la philosophie académique occidentale, l'idée selon laquelle l'expérience ne fournit aucune substance épistémique qui serait *donnée*, immédiate, et qui pourrait servir de fondement absolu à la connaissance. Au contraire, toute expérience est modelée par la culture dans laquelle un agent ou une agente épistémique baigne; en sciences, cela se traduit notamment par la charge théorique des observations.

Ce rejet du fondationnalisme doit être distingué d'un rejet de l'empirisme puisque les auteurs et autrices ayant développé des approches cohérentistes ont conservé l'idée cruciale selon laquelle l'expérience joue un rôle dans la détermination des dynamiques scientifiques (une idée à la fois normative et descriptive; voir Longino 1990). Ce conservatisme est parfaitement illustré par les réponses que les auteurs contemporains et autrices contemporaines ont données au problème de la coordination (voir sections 2.3.1 à 2.3.3 du présent chapitre). Bien que ce problème souligne l'inévitabilité de la charge théorique des observations, il place aussi ces observations empiriques, leur standardisation et leur progrès au cœur des dynamiques scientifiques. Le défi pour les approches cohérentistes est donc de rendre compte de ce rôle que joue l'expérience, sans pour autant tomber dans les écueils du fondationnalisme. Avant d'analyser en détail les approches de Chang, de van Fraassen et de Tal qui répondent avec brio à cette injonction, il est important d'illustrer les pièges propres au cohérentisme qu'il est prudent d'éviter. Pour ce faire, je fais une lecture naïve du constructivisme (voir Latour 2003 pour un aperçu de la différence entre un constructivisme naïf et une version rigoureuse qui s'apparente au cohérentisme historique défendu dans cette thèse), qui est l'une des approches spécifiques associées au cohérentisme. L'objectif ici consiste à mettre l'accent sur certaines difficultés potentielles que des réponses cohérentistes au problème de la coordination peuvent engendrer pour une philosophie des sciences centrée sur l'expérience.

Le constructivisme, représenté par les écrits de personnes comme Latour (1987), Haslanger (1995), Schaffer (1992) ou Hacking (1999), réfère généralement à la posture

philosophique selon laquelle les connaissances (scientifiques et autres) sont des produits (ou construits) sociaux. En philosophie de la mesure, cela revient notamment à stipuler que la régularité des phénomènes que l'on cherche à mesurer est un artéfact résultant des efforts de contrôle humain et de conditions d'observation bien particulières. Tal exprime la différence entre cette perspective et le conventionnalisme comme suit : « According to constructivists, standardizing time is not a matter of choosing which pre-existing natural regularity to exploit; rather, it is a matter of constructing regularities from otherwise irregular instruments and human practices. » (Tal 2016, 314). Tal mobilise aussi un exemple, soit la standardisation des unités de temps, pour illustrer cette distinction.

Comme mentionné précédemment, de concert avec Reichenbach, il est impossible de juxtaposer deux périodes successives d'une horloge; il est donc impossible de démontrer sa régularité, c'est-à-dire l'égalité de ses périodes. Cette régularité, en pratique, est établie en comparant plusieurs horloges et en utilisant des modèles qui permettent d'analyser les différences entre les horloges et de faire des diagnostics tels que « cette horloge est moins régulière que l'autre, puisque son fonctionnement s'écarte plus de la norme prédite par le modèle ». C'est ainsi que les différentes horloges sont utilisées par les métrologues contemporains afin de réguler le Temps Universel Coordonné : les anomalies ou différences qui surviennent inmanquablement sont corrigées par l'entremise d'un modèle prédictif (Tal 2016). En ce sens, l'établissement d'un standard dépend de lois postulées, celles-ci n'étant évidemment vérifiables que par l'entremise des horloges en question (elles-mêmes fabriquées sur la base d'éléments théoriques partagés avec le modèle). Aussi bien dire que ce qui est considéré, aujourd'hui comme par le passé, comme la seconde universelle ne détient aucun corrélat physique isolé. C'est plutôt le fruit d'une mise en commun de plusieurs standards physiques à partir desquels l'extraction d'un point de référence international nécessite un appareillage théorique et matériel massif (Tal 2016). La régularité du temps, tout comme son standard de mesure, est une *construction* (Latour 1987).

Cette conception des objets scientifiques comme étant des constructions s'oppose à celles qui les considèrent comme des réalités ontologiques. Cette posture réaliste est associée par Tal à ce qu'il appelle la thèse de la réification (Tal 2016). Précisément, cette thèse de la réification

consiste à affirmer que des objets théoriques, *parce qu'ils ont un corrélat physique*, peuvent être considérés comme des choses (*res*) réelles détenant un statut ontologique allant bien au-delà de leur statut d'artéfact. Or, comme illustré précédemment, cette manière de procéder nie la complexité des objets théoriques inscrite au cœur des pratiques scientifiques. La seconde universelle, par exemple, est une abstraction atteinte par l'interaction entre une multitude d'horloges et de modèles prédictifs, ce qui rend impossible sa localisation physique. Tout corrélat physique de la seconde universelle est donc soit approximatif, soit illusoire.

En identifiant et en s'opposant à ce phénomène de réification commun en sciences et chez les philosophes adeptes de postures fondationnalistes, le constructivisme opère du même coup un renversement d'explanandum et d'explanans. Reichenbach, par exemple, considère que la régularité d'un phénomène, acceptée en tant que principe définitionnel, peut être mobilisée comme fondement et explication des pratiques métrologiques, de leurs succès et de leurs transformations. Les constructivistes suggèrent au contraire que cette régularité, loin d'être une chose qui explique le succès des pratiques métrologiques, est le fruit des pratiques scientifiques, l'explanandum dont on peut comprendre l'apparition et les modalités par l'entremise de l'étude des dynamiques des pratiques scientifiques.

La standardisation de pratiques scientifiques, comme celles liées au temps et à la seconde universelle, doit donc être perçue comme un acquis, une réalisation cruciale toujours ancrée dans un dynamisme historique, un aboutissement (temporaire) que philosophes, sociologues et anthropologues des sciences se doivent d'étudier, de décrire et idéalement d'expliquer. De fait, pour rendre compte adéquatement des pratiques métrologiques, il faut donc s'attarder à ce qui rend possible ce genre de standardisation, ce qui sous-tend les succès métrologiques. Selon le constructivisme *naïf*, cela implique de s'intéresser d'abord et avant tout aux dynamiques sociales de la communauté scientifique concernée et de délaissier les autres aspects de la structure épistémique (notamment la théorie, sa structure logique et son rapport à l'observation).

Une difficulté rencontrée par cette version du constructivisme est de pouvoir expliquer de manière satisfaisante la stabilité de certaines pratiques de mesure et, plus particulièrement, du plus grand succès de certaines de ces pratiques par rapport à d'autres. Tal, toujours en discutant

de la standardisation des unités temporelles, utilise le concept de stabilité pour référer au besoin d'ajustements moins prononcé de certains systèmes, une caractéristique considérée comme étant souhaitable. Selon cet auteur, il est indéniable que certains systèmes de standardisation soient plus stables que d'autres :

To illustrate this point, imagine that metrologists decided to keep the same algorithm they currently use for calculating UTC [Universal Time Coordinated], but implemented it on the human pulse as a standard clock instead of the atomic standard. As different humans have different pulse rates depending on the person and circumstances, the time difference between these organic standards would grow rapidly from the time of their latest correction. Institutionally imposed adjustments would only be able to bring universal time into agreement for a short while before discrepancies among different pulse-clocks exploded once more. The same algorithm that produces UTC would be able to minimize adjustments to a few hours per month at best, instead of a few nanoseconds when implemented with atomic standards. (Tal 2016, 317)

La différence de stabilité entre deux systèmes de standardisation évoquée par cet exemple se doit d'être expliquée par une approche philosophique des pratiques métrologiques.

Le conventionnalisme, en tant qu'approche fondationnaliste, expliquerait cette plus grande stabilité du système métrologique mobilisant l'horloge atomique par la plus grande régularité du phénomène sous-jacent au fonctionnement de cette horloge. Comment peut-on savoir si cette explication est valide? Puisque la coordination implique des pétitions de principe, on ne peut pas le garantir, mais le conventionnalisme transforme alors la régularité des horloges (du phénomène sous-jacent) en vérité définitionnelle, en affirmation sémantique sur laquelle repose la connaissance. Cela dit, ce geste épistémologique permet de défendre la stabilité de toute horloge et est donc insuffisant pour distinguer les horloges cardiovasculaires et atomiques mentionnées par Tal (2016). Or, une fois que les deux systèmes sont ainsi représentés comme étant fondamentalement valables, il ne reste plus qu'à choisir le plus efficace, soit celui mobilisant l'horloge atomique.

Bien que cette logique soit rassurante, elle est lacunaire. Comme démontré plus tôt, elle prétend d'une part que la sémantique est extérieure à la connaissance (sans quoi la circularité que le conventionnalisme cherche à éviter serait maintenue). D'autre part, elle explique la validité d'un phénomène humain et social, soit la régularité d'un système épistémique, par des faits

inaccessibles, soit la régularité de phénomènes. La question est maintenant de savoir si le constructivisme peut faire mieux.

Pour ce faire, l'approche constructiviste naïve doit mobiliser un explanans autre que celui dénoncé par l'entremise de la thèse de la réification. Elle se tourne plutôt vers les dynamiques socioculturelles menant à la standardisation. Le problème, comme souligné par Tal, est que la communauté métrologique pourrait appliquer les mêmes méthodes et opérations, soutenues par le même réseau d'agents et d'agentes épistémiques, mais obtenir des résultats différents selon le type d'horloge utilisé. Le système mobilisant l'horloge atomique est plus stable que celui faisant appel à l'horloge cardiovasculaire; il nécessite donc moins de corrections. Le même traitement social (l'application du même nombre de corrections; il s'agit bien sûr ici d'une simplification heuristique) des deux phénomènes (le pouls et le changement de niveau d'un électron) entraîne des résultats métrologiques différents. Conséquemment, l'une des deux avenues est, dans la pratique (qui est notre objet d'étude, rappelons-le), considérée comme étant supérieure à l'autre, un fait qui doit être expliqué par toute approche épistémique cohérentiste qui prétend décrire les dynamiques scientifiques de manière satisfaisante. En effet, si une version du cohérentisme, comme le constructivisme naïf centré sur les relations sociales entre chercheurs et chercheuses, est incapable de rendre compte de la supériorité de certains systèmes épistémiques face à d'autres, elle est alors de fait incompatible avec l'aspiration des sciences à la normativité : la révision, le rejet et le dépassement de systèmes épistémiques sont synonymes de progrès scientifique.

2.3 Le cohérentisme historique

Le développement des approches philosophiques du problème de la coordination, au 20^e siècle, nous laisse dans une situation peu enviable : d'un côté, le fondationnalisme (par l'entremise du conventionnalisme) semble avoir échoué dans sa tentative de rendre compte de la structure épistémique des sciences; de l'autre, le constructivisme est mal équipé pour conceptualiser la dimension normative des pratiques scientifiques. C'est principalement pour remédier à cette impasse épistémique que des auteurs et autrices du 21^e siècle ont repensé les conséquences philosophiques du problème de la coordination en mettant l'accent sur les

dynamiques diachroniques qui sous-tendent la justification empirique des théories et énoncés scientifiques. J'utilise l'étiquette « cohérentisme historique » pour référer à ce courant de pensée en bloc.

Dans ce qui suit, je m'attarde plus spécifiquement sur les travaux de Chang, de van Fraassen et de Tal en mettant l'accent sur les points saillants de leurs approches qui peuvent être importés au sein d'une analyse philosophique de la phylogénétique centrée sur le problème de la coordination. Leurs approches et d'autres approches similaires (Frigerio, Giordani et Mari 2010; Israel-Jost 2019; Mari 2005; Morrison 2009; Shapin et Schaffer 1985) témoignent de la mise en place d'un quasi-consensus en philosophie de la mesure selon lequel il faut trouver une alternative viable au fondationnalisme, un point de vue partagé par certains praticiens et certaines praticiennes :

Several scientists, significantly coming from different disciplines (for example, although with standpoints and conclusions not fully coincident, Bateson, Morin, and Varela), are now proposing a 'non fundamental' standpoint, according to which no absolute foundation is possible for science, because human knowledge is essentially based on a continuously iterative, try-and-revise, adaptive, autopoietic process [...] in which progressively some elements become more and more solid but nothing is definitive. In this view, knowledge [is] not a building to be founded, but network of components sustaining each other and assuming a meaning only in the context they are contributing to create. (Mari 2005, 265)

Malgré la dominance actuelle de cette posture au sein de la philosophie et de l'histoire de la mesure, je réserve l'espace nécessaire pour analyser une objection qui a récemment été formulée à l'égard du cohérentisme revisité (Isaac 2019). La critique d'Isaac me permet notamment de présenter les approches réalistes en philosophie de la mesure que j'ai jusqu'à présent contournées (parce qu'elles négligent le problème de la coordination et ses conséquences). L'argument d'Isaac se distingue de la posture réaliste traditionnelle en visant explicitement à rendre compte des boucles épistémiques associées à la mesure et à les réconcilier avec une posture réaliste.

2.3.1 Hasok Chang : itérations épistémiques et cohérentisme progressif

Chang suggère que le problème de la coordination et, de manière générale, la charge théorique des observations empiriques nous force à adopter un point de vue cohérentiste

lorsqu'il est question de décrire la structure épistémique de pratiques scientifiques. Chang adopte une approche qu'il nomme le cohérentisme *progressif*. Celle-ci met l'accent sur le processus itératif au cœur du dynamisme des sciences en fournissant des concepts qui permettent de penser le progrès et la normativité. Contrairement au constructivisme naïf, l'approche de Chang permet d'inclure les dynamiques sociohistoriques des communautés scientifiques dans une analyse épistémologique sans pour autant réduire l'épistémologie à l'ethnologie ou à l'histoire des sciences.

Au cœur du cohérentisme progressif de Chang, on retrouve le processus d'itération épistémique que le philosophe et historien des sciences considère comme une notion élargie d'itération (*a broadened notion of iteration*) telle qu'elle est comprise et utilisée en mathématique (et en informatique) :

Epistemic iteration is a process in which successive stages of knowledge, each building on the preceding one, are created in order to enhance the achievement of certain epistemic goals... In each step, the later stage is based on the earlier stage, but cannot be deduced from it in any straightforward sense. (Chang 2004, 226)

Tout comme dans le cas de l'itération mathématique, chaque étape du raisonnement est reconnue comme étant approximative ou imparfaite, mais tout de même porteuse d'une amélioration. La différence principale entre l'itération mathématique (processus qui est à la base de l'apprentissage automatique [*machine learning*]) et l'itération que Chang associe aux activités épistémiques en général est le but vers lequel chacune tend.

L'itération mathématique vise le développement d'algorithmes ou de modèles. Ceux-ci sont élaborés en utilisant des résultats déjà connus ou connaissables par d'autres moyens. En phylogénétique, par exemple, ces processus d'itération permettent notamment de développer des algorithmes pour la reconstruction d'arbres. Un arbre phylogénétique est d'abord postulé pour servir de cible¹², l'objectif étant de développer des méthodes qui seraient capables de le reconstruire à partir de données partielles (par exemple, à partir de données portant sur les liens de similarité entre des unités taxonomiques opérationnelles). Un modèle de base peut alors être

¹² Le processus d'itération peut aussi utiliser des données réelles comme cibles lorsque celles-ci sont disponibles.

mis en place. Ce modèle traite l'information disponible et mène à la reconstruction d'un arbre phylogénétique. Cette démarche est par la suite réitérée en apportant des changements au modèle de base (par exemple, en donnant un poids différent à certains traits utilisés dans l'analyse). Les résultats de ces itérations successives peuvent être évalués sur la base de leur correspondance à l'arbre phylogénétique postulé comme cible, de sorte à identifier les meilleures combinaisons de paramètres. Le champ des possibles est ainsi exploré. Au besoin (si la puissance calculatoire est trop grande pour que toutes les options soient évaluées), une heuristique peut être mobilisée pour guider les itérations successives dans les directions les plus prometteuses. Le modèle et les paramètres qui engendrent la reconstruction la plus proche de l'arbre fictif peuvent par la suite être utilisés pour le traitement des données pour lesquelles la phylogénie est inconnue, c'est-à-dire des données issues du monde biologique. À ce stade, cependant, le processus itératif a déjà accompli sa fonction, soit celle d'identifier des paramètres adéquats pour la reconstruction d'arbres utilisant un modèle donné.

Le processus d'itération épistémique fonctionne sans évaluer les pratiques par rapport à un résultat déjà connu qu'elles devraient obtenir. Le développement de la thermométrie, par exemple, peut être considéré en rétrospective comme tendant vers un point donné (les connaissances contemporaines) qui pourrait servir de point de référence pour l'évaluation des itérations préalables. Il s'agit cependant d'un jeu de perspectives trompeur (les travaux de van Fraassen s'intéressent précisément à ce genre de jeu de perspectives; voir section 2.3.2). Comme Chang le souligne, les agents et agentes épistémiques impliqué·e·s à un moment donné de l'histoire de la thermométrie ignorent le point vers lequel tend la succession des itérations constitutives de l'ensemble de pratiques qu'on appelle la thermométrie. La validité des solutions proposées à un problème donné (par exemple, l'établissement de la fonction reliant température, pression et volume) est évaluée à l'aune de critères variés, établis notamment par les itérations passées, sans faire appel à la solution ultime. Chang souligne également qu'il est impossible de déduire un état de connaissance (*state of knowledge*) à partir de l'état préalable puisqu'un grand nombre de facteurs variés viennent déterminer les changements opérés (facteurs sociohistoriques, importations de principes ou de méthodes issus d'autres domaines scientifiques, innovations techniques, etc.).

Pour rendre la notion d'itération utile à l'épistémologie, il faut bien saisir qu'elle s'applique à différents niveaux d'analyse. Dans une perspective englobante, on peut parler de différentes itérations successives d'un complexe discursif, d'un ensemble large de pratiques scientifiques. En thermométrie, pour reprendre un exemple de Chang, l'utilisation généralisée du thermomètre à air peut être considérée comme une itération thermométrique ayant suivi l'utilisation généralisée du thermomètre au mercure (Chang 2004). En ce qui concerne la classification biologique, la classification phylogénétique peut être considérée comme une itération de la biologie comparative ayant suivi les classifications prédarwiniennes, préphylogénétique (bien que d'autres itérations non phylogénétiques aient continué à se développer en parallèle). D'une itération à l'autre, des changements sont induits dans le système épistémique, changements qui, pour avoir un impact à cette échelle, doivent être significatifs. La réinterprétation darwinienne et phylogénétique des systèmes classificatoires en est un bon exemple.

Dans une perspective plus locale, l'itération épistémique réfère à la succession de pratiques bien déterminées et circonscrites. Par exemple, un ou une scientifique utilise les résultats d'une expérience passée pour peaufiner son protocole, de sorte que la nouvelle itération de l'expérience sera différente, mais indéniablement en continuité avec celle qui l'a précédée. En phylogénétique, l'utilisation de groupes extérieurs pour résorber un conflit de congruence entre différentes homologues postulées s'inscrit dans un tel processus d'itération. En effet, on utilise alors des résultats de pratiques classificatoires passées pour ajouter de l'information à nos systèmes ou pour les réviser, sans transformer notre méthode de manière significative.

Dans ces deux situations, celle plus locale et celle plus générale, le modèle de Chang vise à décrire les dynamiques scientifiques telles qu'il les a observées en portant son attention sur (et en guidant la nôtre vers) différents épisodes de l'histoire des sciences (Chang 2004, 2012, 2015). En effet, peu importe à quelle échelle on observe ces dynamiques, force est d'admettre que la structure justificative d'un état de connaissance (qu'il soit traduit par un article, un livre, les résultats d'une expérience, la présentation d'un protocole, la synthèse des connaissances d'un domaine, un rapport, un texte de vulgarisation, etc.) dépend des itérations passées. Ce phénomène s'explique par l'omniprésence du problème de la coordination : le va-et-vient justificatif reliant théorie et observations implique l'impossibilité de considérer les phénomènes,

le noumène, l'expérience, le donné, les données ou encore le monde « réel » comme la source de laquelle les pratiques scientifiques tirent leur légitimité. Si nous remarquons que les travaux scientifiques s'appuient sur des travaux passés, c'est bien parce que les scientifiques doivent constamment gérer les difficultés de la coordination sans pour autant abandonner l'empirisme (il est fort possible qu'une dynamique similaire se mettrait en place au sein de complexes discursifs non empiriques, mais une telle étude dépasse le cadre de cette thèse). En d'autres termes, puisqu'il y a problème de la coordination, il est attendu que la structure épistémique d'un complexe discursif mobilisant l'expérience détienne une composante diachronique que Chang suggère de penser par le modèle de l'itération épistémique.

Le moteur du changement, dans ce modèle¹³, est ce que Chang appelle l'impératif du progrès (*imperative of progress*) : chaque itération s'appuie sur les précédentes, mais cherche à faire mieux. « Faire mieux », ici comme ailleurs dans cet ouvrage (et dans les travaux de Chang), est à penser à l'aune du problème de la coordination, c'est-à-dire avec l'impossibilité d'une fondation absolue à la connaissance. Autrement dit, « faire mieux », c'est tendre vers la satisfaction d'une variété de valeurs épistémiques considérées comme étant appropriées par les praticiens et les praticiennes ainsi que la communauté qui évalue leur travail (philosophie, histoire et anthropologie des sciences, institutions chargées du financement et de l'éthique de la recherche, société civile, etc.). Chang mentionne quelques valeurs épistémiques typiquement reconnues comme étant souhaitables en s'inspirant des grands noms de la philosophie des sciences : la simplicité d'une itération, son imbrication au sein d'un système épistémique plus grand, sa portée, son élégance, sa puissance unificatrice ou explicative, etc.

En ce sens, Chang préconise une attitude pluraliste, et de surcroît, pragmatiste. En effet, il insiste sur le fait qu'il y a une diversité de manières par lesquelles un état de connaissance donné peut être considéré comme une avancée et surtout qu'aucune de ces modalités ne doit être priorisée face aux autres. Tant que les valeurs d'évaluation acquièrent une légitimité au sein de

¹³ Pour plus de détails sur l'itération épistémique et le principe de respect (mentionné plus loin), voir Chang 2004, 44-46 et chapitre 5, ainsi que Chang 2007. Pour plus d'informations au sujet de l'impératif du progrès, voir Chang 2004, chapitre 5.

la communauté scientifique, et idéalement au sein de la culture intellectuelle dans laquelle les sciences sont pratiquées, alors elles sont dignes de mention et d'utilisation. Une telle perspective est cohérente avec les approches qui mettent l'accent sur l'effet des valeurs non épistémiques (sociales et politiques) sur l'évaluation des théories et hypothèses scientifiques en plus de considérer l'importance de multiples valeurs épistémiques (Brigandt 2020; Longino 1990; Elliott 2011, 2017).

Selon Chang, l'attrait de l'utilisation d'une pluralité de critères normatifs s'explique notamment par le fait que la notion de vérité s'avère inapte à l'évaluation des pratiques scientifiques. On peut en effet se poser la question suivante : comment pourrait-on évaluer la validité d'une procédure métrologique à l'aide de la notion de vérité? Une réponse possible serait de considérer la convergence de différentes méthodes comme une manière adéquate d'établir la « vraie valeur » de la température d'un phénomène. Différentes procédures pourraient ensuite être évaluées grâce à cette « vraie valeur. » La « vérité » serait cependant complètement soumise au processus d'itération épistémique (qui, comme démontré plus tôt, découle du problème de la coordination). C'est seulement grâce au processus itératif que les méthodes d'observation peuvent être perfectionnées pour éventuellement converger (il est en effet plutôt compliqué d'établir la correspondance de la hauteur d'une colonne de mercure avec celle d'une colonne d'alcool, du moins en termes de température, sans référer aux *acquis* de la thermométrie). Ce processus requiert cependant des critères de validité pour progresser adéquatement et rendre possible la convergence de résultats mobilisant des méthodes et appareils d'observations drastiquement différents. En ce sens, la vérité est le résultat d'un processus itératif et ne peut donc pas servir de critère pour l'évaluation des différentes itérations (ici, il est primordial de garder en tête la différence entre l'itération épistémique et l'itération mathématique mentionnée plus tôt). Chang résume la chose comme suit :

In case we have a convergent operationalization, we could consider the limits of convergence as the "real" values; then we can use these values as the criteria by which we judge whether other proposed operationalizations produce true values. But we must keep firmly in mind that the existence of such "real values" hinges on the success of the iterative procedure, and the successful operationalization is constitutive of the "reality." If we want to please ourselves by saying that we can approach true values by

iterative operationalization, we also have to remember that this truth is a destination that is only created by the approach itself. (Chang 2004, 217)

De ce fait, Chang conclut qu'il vaut mieux se tourner vers la variété de critères mentionnés précédemment, dans une perspective pluraliste. Au-delà de cette prise en compte rigoureuse des conséquences du problème de la coordination, il décrit les dynamiques scientifiques de sorte à bien démontrer que la circularité mise à jour par les auteurs et autrices classiques (en lien avec le problème de la coordination) est illusoire : le rapport justificatif entre procédures d'observation et théories est mutuel et superficiellement circulaire, mais l'ensemble mobilise ultimement quelque chose qui lui est extérieur, soit les itérations passées. Celles-ci peuvent difficilement être considérées comme des *bases fondamentales* puisqu'elles aussi sont imprégnées du problème de la coordination (elles s'appuient donc elles aussi sur des itérations les précédant), mais ce sont tout de même des bases à la connaissance ou, du moins, *des sources de légitimité*. En puisant dans les itérations passées, les systèmes de connaissances et leur progrès évitent ainsi les charges de relativisme débridé, malgré leur teneur éminemment historique et l'acceptation du pluralisme normatif décrit précédemment. De surcroît, cette perspective souligne de manière efficace l'aspect processuel des pratiques scientifiques (Hull 1988).

Afin de rendre compte du fait que les états de connaissance puisent inmanquablement dans les itérations passées tout en cherchant à les améliorer, Chang théorise la notion de *principe de respect* : s'il faut éviter d'attribuer aux standards procéduraux passés une autorité absolue, ce qui rendrait impossible tout progrès, il est cependant tout à fait acceptable et souhaitable de reconnaître la validité (provisoire) et l'utilité de résultats préalables. Ainsi, on peut témoigner d'un respect pour un ensemble de procédures, tout en cherchant à les améliorer. Selon le modèle de Chang caractérisé par l'impératif du progrès conjugué au principe de respect, la circularité justificative laisse place à une spirale : « Progress comes to mean a spiral of self-improvement if it is achieved while observing the principle of respect » (Chang 2004, 44). Il va sans dire que cette approche a grandement influencé la description de la cladistique hennigienne et la brève histoire de l'homologie présentées au chapitre précédent, deux situations empiriques qui viennent du même coup renforcer la validité du modèle proposé pour Chang au-delà de la métrologie. Je préfère cependant parler d'un modèle hélicoïdal plutôt que d'un modèle basé sur la spirale

puisque cela permet de se représenter plus facilement la dimension temporelle des dynamiques scientifiques (voir Figure 1.4).

Afin de compléter la description du modèle de Chang, il faut s'attarder à la notion de principe ontologique que l'auteur introduit au second chapitre de son ouvrage de 2004. Lors d'une expérience dont les résultats ont été publiés en 1847, Regnault (1810-1870) teste l'efficacité des deux types de thermomètres, non pas sur la base de leur conformité aux prédictions d'une théorie donnée (la loi des gaz parfaits avait déjà été formulée à l'époque), mais bien sur la base de leur comparabilité, soit leur capacité à reproduire les mêmes résultats à plusieurs reprises dans les mêmes conditions. Ainsi, il compare différents thermomètres au mercure, et ce, à de nombreux points sur l'échelle centigrade, puis il impose le même régime au thermomètre à air. À la suite de cette expérience, bien qu'il ait noté la comparabilité faillible des thermomètres à air, il conclut que ceux-ci sont tout de même plus fiables que ceux au mercure, une conclusion à laquelle la communauté scientifique se rallie rapidement.

La question qu'il faut se poser ici est la suivante : pourquoi la comparabilité est-elle un critère d'évaluation valable pour guider le progrès scientifique? Pour Regnault, l'objectif explicite était de minimiser les prémisses théoriques nécessaires pour concevoir un test qui départagerait les deux types de thermomètres (Chang 2004). Cependant, la comparabilité elle-même apparaît comme une prémisse théorique qui fait resurgir de plus belle le problème de la coordination. En effet, on postule que lorsque les conditions matérielles sont (approximativement) les mêmes, un thermomètre devrait toujours indiquer la même température. Sur cette base, on tente d'établir le type de thermomètre qui est le meilleur. Or, pour confirmer la validité de ladite prémisse, il faudrait évidemment être muni de thermomètres dont on connaît de manière indubitable la validité. La comparabilité n'amène donc aucune fondation à la connaissance qui serait plus stable que celle fournie dans une logique relative au principe de respect décrit plus tôt.

Pour Chang, la comparabilité a toutefois un statut particulier puisqu'elle s'appuie sur ce qu'il appelle un principe ontologique, soit le principe de valeur unique. Un principe ontologique est une croyance détenant une importance critique pour un système épistémique. Ces prémisses sont habituellement injustifiées, voire injustifiables, mais servent tout de même de gonds

(Wittgenstein 1969) sur lesquels s'articulent les raisonnements scientifiques. Chang définit ces principes comme suit : « Ontological principles are those assumptions that are commonly regarded as essential features of reality within an epistemic community, which form the basis of intelligibility in any account of reality. » (Chang 2004, 91). Il précise de plus que ces principes sont bien des croyances qui ne sont pas empiriquement démontrées ni logiquement déduites, de sorte qu'ils peuvent être associés aux gonds (*hinges*) de Wittgenstein (1969). Le rôle de ces principes au sein des structures justificatives pousse Chang à adopter une posture pragmatiste concernant la justification, refusant d'imposer à celle-ci le besoin d'être une justification *absolue* pour être une justification valide (voir Chang 2004, p. 42 pour une référence explicite à Wittgenstein).

Les croyances décrites par Chang comme étant des principes ontologiques pourraient aussi être considérées comme des conventions par Reichenbach ou ses prédécesseurs. En revanche, les deux notions renvoient à des rôles bien distincts que des croyances peuvent occuper au sein d'un modèle des dynamiques des pratiques scientifiques. Pour Chang, ces croyances font partie de la toile qui octroie un sens et une intelligibilité aux pratiques, aux itérations épistémiques. Elles sont privilégiées en ce sens qu'elles sont rarement remises en question par la pratique afin de favoriser le dynamisme mais l'analyse épistémologique n'est aucunement engagée de la même façon envers ces principes. Les principes ontologiques sont, d'un point de vue logique, comparables à d'autres croyances de la toile épistémique. Au contraire, pour le conventionnalisme, ces mêmes principes sont extérieurs à la connaissance et lui servent de fondation en décrivant la structure du monde; ils ont en ce sens un statut épistémique privilégié par rapport au reste, surtout d'un point de vue logique.

Revenons à notre exemple de la comparabilité. Ce qui confère à ce critère une validité pour l'évaluation de procédures d'observation, c'est sa cohérence avec un principe ontologique, soit celui de valeur unique. Le principe de valeur unique est l'idée selon laquelle une quantité physique est toujours univoque, c'est-à-dire qu'elle ne présente qu'une seule valeur dans une situation donnée : « a real physical property can have no more than one definite value in a given situation. » (Chang 2004, 90). Si ce principe est accepté, alors on doit en effet s'attendre à ce que deux thermomètres offrent des résultats comparables dans une situation donnée; toute incohérence entre les résultats est attribuable aux procédures et non au phénomène. Cependant,

d'autres croyances sont nécessaires pour conférer à l'expérience de Regnault sa légitimité : il faut considérer que deux types de procédures (thermomètres au mercure et à air) puissent mesurer la même quantité théorique; il faut accepter que le volume, la pression, la quantité et la température d'un corps soient liés; il faut faire preuve de confiance envers les artisans et artisanes qui confectionnent les thermomètres, surtout envers ceux et celles dont les outils n'ont *pas* été testés par Regnault; etc.

L'un des apports importants de l'analyse épistémologique effectuée en suivant les prescriptions de Chang est de rendre explicite la distinction entre deux manières de qualifier le rôle joué par des croyances au sein d'un système de pensée. Du point de vue des praticiens et des praticiennes, certaines croyances, comme les principes ontologiques, jouent un rôle fondamental, ce qui leur octroie une inviolabilité constructive. En effet, en gardant ces idées fixes, il est possible de progresser sur d'autres fronts sans remettre en question l'ensemble du système épistémique à chaque itération. Du point de vue épistémologique, il est important de souligner que toutes les croyances sont égales; elles contribuent toutes à la cohérence du système épistémique qu'elles composent et dont elles sont dépendantes. Cette distinction que je lis dans les travaux de Chang est épistémologiquement très utile puisqu'elle permet de cerner trois niveaux d'analyse distincts : celui sociohistorique qui explique pourquoi et comment certaines croyances en viennent à acquérir une importance marquée pour les scientifiques; celui de l'analyse logique qui s'intéresse au rapport inférentiel entre différentes croyances; celui théorique, qui met de l'avant l'impératif de progrès en sciences.

Bien que Chang ait érigé son approche en appréhendant des pratiques scientifiques reliées à la mesure, il a aussi démontré l'utilité des notions développées pour analyser des pratiques classificatoires (Chang 2012, 2015, 2017). Peu importe l'objet d'étude auquel on l'applique, les points saillants de cette approche demeurent les mêmes. Il s'agit d'une approche informée par le problème de la coordination et ses conséquences épistémiques. Le processus d'itération épistémique offre alors un patron à partir duquel penser les dynamiques historiques, notamment du point de vue des relations logiques entre croyances. La validité de ce modèle a été établie empiriquement et peut guider la description de dynamiques scientifiques au-delà de celles étudiées par Chang. Dans cette optique, l'auteur suggère d'identifier les principes ontologiques

qui viennent structurer les différentes itérations épistémiques (ou états des connaissances) afin de bien cerner les relations entre théories, hypothèses, croyances diverses et observations qui composent le champ à l'étude.

Ce modèle suggère bien sûr que les connaissances sont au moins partiellement construites par des communautés scientifiques qui adoptent des croyances non justifiées (voire non justifiables). Cependant, ce constructivisme (plutôt sage que naïf) évite le relativisme débridé, l'absence de normativité scientifique ou l'impossibilité de caractériser le progrès par l'entremise de valeurs épistémiques dont Chang reconnaît la portée. Celles-ci, jumelées au principe de respect et à l'impératif du progrès, fournissent des repères à partir desquels il est possible de cerner et de commenter la normativité scientifique : pour « faire mieux » que ses prédécesseur·e·s, une itération épistémique doit répondre aux problèmes ou défis posés dans le champ d'études en question en fonction des valeurs préconisées par la communauté concernée. Les réponses aux questions « quels sont les problèmes posés? » et « quelles sont les valeurs à préconiser? » sont situées dans un temps, un espace et un discours donnés. La situation serait (très) différente si les connaissances pouvaient être évaluées sur la base d'une fondation absolue, d'une perspective de « nulle part » ou par l'entremise de révélations divines.

2.3.2 Bastian van Fraassen

Le problème de la coordination reçoit de la part de van Fraassen une attention particulière, comme en témoigne son livre de 2008, *Paradoxes of Perspective*, dans lequel le chapitre 5 y est entièrement consacré. Ce casse-tête épistémologique est utilisé pour structurer les réflexions de l'auteur au sujet de la mesure, réflexions qui animent les deux chapitres subséquents. Dans ce qui suit, je résume les points saillants de l'analyse du problème de la coordination par van Fraassen et la modulation du propos par sa version de l'approche sémantique des théories scientifiques.

2.3.2.1 La vision synoptique

Après avoir considéré les approches traditionnelles au problème de la coordination, van Fraassen suggère que, pour bien gérer les conséquences épistémologiques de ce problème, la

philosophie de la mesure et des sciences gagne à réconcilier deux perspectives complémentaires sur son objet d'étude :

The two ways of looking at the matter we must combine in a synoptic vision. The first is 'from within' the historical process in which measurement procedures and theory are stabilized. The second is 'from above' with the theoretical description of the domain including the measurement interactions already in hand, a stage achieved in that historical process but no longer involving any explicit reference to its own history. (van Fraassen 2008, 139)

Van Fraassen recommande donc la mise en commun d'une perspective interne (*from within*) et d'une perspective surplombante (*from above*). Pour bien comprendre ce que la jonction de ces deux perspectives peut offrir, il convient avant tout de les présenter séparément.

La perspective interne vise à embrasser l'historicité des pratiques scientifiques. Elle est en ce sens semblable à l'approche de Chang. Il s'agit de considérer les différents défis auxquels font face les acteurs et actrices des sciences lorsqu'ils et elles cherchent à standardiser des pratiques de mesure, à peaufiner des protocoles expérimentaux ou des pratiques inférentielles, à établir la validité d'énoncés théoriques, de modèles ou de théories, etc. Toutes ces pratiques impliquant d'une manière ou d'une autre la coordination entre théorie et phénomènes, la stabilisation des états de connaissance requiert nécessairement la mise en place de réponses concrètes au problème de la coordination.

C'est dans une telle optique que Chang a appréhendé les travaux de Regnault (Chang 2004) qui ont confirmé la supériorité du thermomètre à l'air. Dans cette situation, rappelons-le, les phénomènes observés ne peuvent pas trancher le débat puisqu'ils ne sont observables que par l'entremise des outils que l'on cherche à départager, eux-mêmes calibrés sur la base d'éléments théoriques. À l'inverse, la théorie seule est aussi insuffisante pour confirmer la supériorité d'un type de thermomètres puisque sa validité empirique dépend d'outils de mesure et pourrait donc être artéfactuelle, incorporant de fait les erreurs des outils utilisés pour la confirmer. Cette boucle est inévitable, mais les praticiens et praticiennes trouvent immanquablement des manières de négocier leur rapport au monde empirique de sorte à relever les défis épistémiques de leur communauté. C'est ce à quoi je réfère lorsque je mentionne la possibilité de fournir des « réponses concrètes au problème de la coordination ».

Les travaux Regnault en sont un exemple brillant. Il a su mobiliser des croyances disponibles et partagées par l'ensemble de la communauté scientifique (le principe de valeur unique et l'importance de la comparabilité qui en découle) pour réaliser des expériences rigoureuses dont la portée significative s'explique notamment par la réputation d'expérimentateur de haut niveau dont Regnault jouissait déjà. Évidemment, les phénomènes et l'expérience qu'on peut en faire contribuent à l'explication, mais ils font partie d'un ensemble de facteurs variés.

Dans cet exemple, les croyances de la communauté et l'organisation sociale de celle-ci, qui a permis à Regnault d'acquérir la réputation d'expérimentateur hors pair, sont des facteurs essentiels pour comprendre la résolution du conflit épistémique opposant le thermomètre à l'air et celui au mercure. Le tout se résume par la citation suivante : « What now counts as simple passive measurement is a hard-won achievement. » (van Fraassen 2008, 125). En d'autres termes, ce qui semble être un simple fait empirique, quelque chose de *donné* (*given*; Sellars 1956), est en fait le résultat de démarches considérables ayant permis de trancher un débat au sujet de quantités inobservables. La perspective interne met l'accent sur la diversité des facteurs qui déterminent ces démarches et leurs enchaînements (itératifs).

En contraste, la perspective surplombante de van Fraassen consiste à utiliser la stabilité issue des progrès scientifiques pour s'extraire des aléas associés aux processus de standardisation des pratiques. Dans cette optique, un concept théorique est moins à penser en fonction des procédures qui le lient aux phénomènes (et des difficultés qui accompagnent cette liaison) qu'en fonction du rôle qu'il occupe au sein d'une théorie stable et acceptée. En d'autres termes, la validité des procédures de mesure est acceptée *de facto*, leurs résultats sont considérés comme quelque chose de *donné* et on observe les développements scientifiques à travers les lunettes de ladite théorie.

Cette perspective se distingue drastiquement de l'approche proposée par Chang. Il s'agit, d'une certaine manière, de faire vœu d'ahistoricité et de traiter le rapport entre les composantes des pratiques scientifiques à l'aune de la logique prescrite par la théorie stabilisée servant de point d'ancrage. Bien que cette théorie soit issue d'un processus historique, la

perspective surplombante fait abstraction de cette historicité (« *a stage achieved in that historical process but no longer involving any explicit reference to its own history* » [van Fraassen 2008, 139]) et fait régner une logique ne laissant place à aucune ambiguïté.

Pour l'exemplifier simplement, considérons l'expérience de Regnault du point de vue stabilisé de la thermodynamique. De ce point de vue, le thermomètre à air n'est qu'un thermomètre au gaz parmi d'autres, et en ce sens, sa supériorité va de soi. En effet, les caractéristiques de ces thermomètres, prédites par la théorie, sont telles que l'expérience de Regnault n'est qu'une étape dans la démonstration rationnelle de leur supériorité, au même titre que la formulation de la loi des gaz parfaits. Considérant que la dilatation des gaz est stable à toutes températures, que cette dilatation est plus marquée que celle d'autres substances thermométriques et que les gaz sont faciles à isoler, la calibration des thermomètres à air ne peut être que plus facile que celle des autres substances thermométriques disponibles. Autrement dit, les succès de Regnault s'expliquent par les caractéristiques de l'air (et du mercure) telles que les lois de la thermodynamique permettent de les déduire dans un espace abstrait (mathématique). L'enjeu de savoir comment ce complexe logique en vient à acquérir un contenu empirique n'est tout simplement pas au menu. Prise seule, cette perspective réintroduit donc toutes les difficultés associées au conventionnalisme en offrant un regard partiel sur les structures justificatives. Elle offre cependant un cadre au sein duquel l'analyse conceptuelle d'un état de connaissance donné (indépendamment de sa dimension historique) prend tout son sens.

C'est donc dire qu'aucune de ces deux perspectives (interne et surplombante) ne témoigne d'une grande originalité. Cependant, l'injonction de van Fraassen à réunir les deux pour adopter une vision synoptique des dynamiques scientifiques est quant à elle d'un intérêt particulier pour mon projet. Il s'agit de jumeler au sein d'une seule et même analyse l'attention portée à l'historicité d'un complexe discursif et l'analyse logique et anhistorique des liens inférentiels entre les éléments constitutifs d'une itération de ce complexe. En termes compatibles avec l'approche de Chang, cela revient à étudier différentes itérations d'une structure épistémique en ciblant l'historicité du processus, en étudiant ce qui change et ce qui reste stable, une tâche qui requiert d'abord l'analyse de la structure de chaque itération prise indépendamment.

Qu'en est-il de l'étude de pratiques scientifiques contemporaines dont le contexte historique se tisse au moment même où l'on tente de les étudier? La perspective surplombante se fait alors des plus invitantes; il pourrait être adéquat d'étudier, par exemple, des pratiques phylogénétiques d'aujourd'hui en regardant seulement les liens logiques et inférentiels qui les composent, comme en témoignent les articles, livres de référence, ouvrages de vulgarisation et autres documents disponibles. Van Fraassen (ainsi que Chang, Tal et nombre d'auteurs et d'autrices se réclamant d'une version ou l'autre du cohérentisme) nous invite au contraire à replacer cette structure logique dans son contexte sociohistorique afin d'engendrer une compréhension plus fine des dynamiques scientifiques. Il s'agit notamment de mettre à jour les différentes conditions de possibilité du discours d'aujourd'hui. Des liens comme celui entre divergence et sélection naturelle ou comme celui entre phylogénie et similarité peuvent être interprétés comme de telles conditions de possibilité pour un sous-ensemble des pratiques phylogénétiques d'aujourd'hui. Or, ces liens inférentiels qui semblent nécessaires et évidents, à peine dignes de mention (perspective surplombante), sont aussi des liens inférentiels construits sociohistoriquement pour lesquels aucune justification absolue n'existe (perspective interne). Ces liens entre croyances n'en demeurent pas moins les seuls acquis permettant d'éviter que la recherche soit paralysée par les difficultés liées à la coordination et les objets d'étude privilégiés d'une épistémologie appliquée laissant place à l'analyse critique consciente de l'histoire des croyances en jeu (vision synoptique).

L'injonction, nécessaire pour les pratiques scientifiques, à traiter certaines croyances avec révérence, à les accepter de fait, est donc délaissée par l'épistémologie adoptant une vision synoptique : toutes les croyances sont mises sur un pied d'égalité et il s'agit simplement de comprendre le comment et le pourquoi des relations qu'elles entretiennent. Ce comment et ce pourquoi détiennent des dimensions historique et construite, certes, mais ils détiennent aussi des dimensions inférentielle, logique et *empirique* : les melons d'eau sans pépins sont *construits* ou créés par des agents et des agentes épistémiques, mais ils sont construits à partir d'une matière tangible, concrète et organique, soumise à diverses restrictions et contraintes (Chang 2004). L'ensemble des déterminants doit être considéré pour comprendre le processus.

2.3.2.2 Van Fraassen et l'approche sémantique des théories scientifiques

La philosophie de la mesure de van Fraassen se caractérise aussi par son ancrage au sein de l'approche sémantique des théories scientifiques. Selon l'approche sémantique, les théories sont constituées de structures épistémiques moins englobantes, soit des modèles. Ces modèles sont composés de relations entre variables, chaque variable pouvant, du moins en théorie, se voir attribuer une valeur. Les phénomènes sont, de la même façon, composés de relations entre différents processus ou aspects physiques (comme le volume, la température et la pression, ou du moins leurs corrélats physiques).

Selon la version de l'approche sémantique de van Fraassen, une théorie ou un modèle est considéré comme étant empiriquement adéquat s'il *sauve les phénomènes* (*if it saves the phenomena*) en les représentant sous une forme qui leur serait isomorphe. L'isomorphie concerne la relation entre la structure des modèles, soit les relations entre leurs composantes et celles des phénomènes. Par exemple, la théorie de l'évolution par voie de sélection naturelle requiert un modèle de l'hérédité qui, peu importe la forme qu'il prend, implique nécessairement une relation entre similarité et phylogénie. Ce modèle est considéré comme étant empiriquement adéquat si et seulement si la même relation est observée (directement ou indirectement) au sein des phénomènes. Si c'est le cas, le modèle convient aux phénomènes, il les « sauve » d'autres modèles qui, au contraire, les trahiraient.

La mesure, ou toute pratique descriptive comparable, rend la structure des phénomènes observable¹⁴, du moins indirectement. Par exemple, la colonne de mercure atteint une hauteur observable dans un thermomètre, ce qui nous donne accès indirectement à la température d'un phénomène. Grâce aux observations ainsi faites, on peut évaluer l'isomorphie de la relation volume-température, telle qu'elle apparaît au sein du modèle, avec la relation des aspects du phénomène correspondant à ces deux quantités théoriques (c'est-à-dire qu'on peut tester les prédictions du modèle). À ce titre, la mesure est un rouage incontournable des sciences

¹⁴ Van Fraassen n'utiliserait pas ce terme dans ce contexte-ci, mais je me permets de le faire pour simplifier l'exposition de ses idées phares qui informent la compréhension du problème de la coordination.

empiriques : c'est la condition de possibilité de la corroboration de la relation d'isomorphie puisqu'elle coordonne des quantités théoriques avec des aspects des phénomènes.

Les sciences existeraient-elles s'il était possible d'avoir un accès épistémique direct aux phénomènes? Aurions-nous besoin de théories et d'outils d'observation pour les décrire, les comprendre et les manipuler? Il est fort probable que la réponse à ces questions spéculatives soit négative. Si, au contraire, l'accès direct aux phénomènes est impossible, alors les sciences prennent tout leur sens, certes, mais elles sont nécessairement accompagnées du problème de la coordination. S'il y a problème de la coordination, alors les approches philosophiques et épistémologiques de la connaissance gagnent à en tenir compte. L'approche sémantique offre une manière de gérer ce problème en traitant la coordination comme l'établissement d'une isomorphie (entre relations composant le phénomène et relations composant le modèle qui le représente) que la mesure permet d'évaluer sans pour autant transformer l'expérience en fondement absolu.

Nonobstant l'attrait de cette approche qui vient donner un cadre philosophique rigoureux pour appréhender le problème de la coordination, elle comporte des limites importantes et critiquables. Une de ces critiques est exposée dans la section suivante (2.3.3). Cette critique ne m'empêche aucunement de me réapproprier la vision synoptique suggérée par van Fraassen; le mariage entre perspectives interne et surplombante permet de tenir compte de l'historicité d'un système de croyances et de pratiques tout en explorant la structure interne et logique du système étudié.

2.3.3 Approches centrées sur les modèles

Les activités de modélisation sont de plus en plus communément reconnues comme détenant un rôle pivot pour les activités de mesure (Frigerio, Giordani et Mari 2010; Mari 2005; Mari et Giordani 2012; Morrison 2009; Morrison et Morgan 1999; Tal 2017). De fait, une famille d'approches centrées sur les modèles visant à la fois à répondre aux questions fondamentales de l'épistémologie et à celles liées à l'utilisation pratique de modèles métrologiques a émergé en philosophie de la mesure. Ces approches sont compatibles avec les grandes lignes de celles de Chang et de van Fraassen, mais elles offrent une cible bien délimitée pour la description des

dynamiques scientifiques en focalisant le propos sur les différents aspects de l'utilisation de modèles par les métrologues. Les approches centrées sur les modèles mettent ainsi l'accent sur la dimension théorique des pratiques de mesures, quelque peu négligée bien que reconnue par le cohérentisme progressif de Chang. Dans ce qui suit, j'illustre ces approches en résumant celle de Tal (2012, 2016, 2017), qui a été retenue en raison du rôle central qu'y joue le problème de la coordination.

Il convient d'abord et avant tout de définir la notion de modèle, non pas de manière à trancher tous les débats sur la question, mais simplement pour rendre le propos intelligible. Tal se rapporte à ce sujet aux travaux de Morrison (2009) et de Cartwright (1999). Selon ces autrices et Tal, un modèle est une représentation abstraite d'un phénomène qui favorise son explication ou les prédictions faites à son égard. Les modèles sont réalisés sur la base de diverses prémisses (*assumptions*) concernant les phénomènes ciblés. Ces prémisses peuvent être issues d'une théorie déjà formulée, d'un traitement préalable de données ou encore de croyances spécifiques (et parfois *ad hoc*, nous dit Tal) qui permettent de faire abstraction de certains aspects plus complexes du phénomène visé. De plus, il est important de noter que les modèles, bien qu'ils soient construits à l'aide d'éléments issus de théories, fonctionnent indépendamment de ces théories (une fois qu'ils sont construits). On peut démontrer qu'un modèle est inadéquat sans que cela n'affecte la théorie, ou vice versa. Cette indépendance explique pourquoi et comment les modèles peuvent être considérés par un grand nombre de philosophes des sciences comme étant à l'interface des aspects symboliques des sciences et du monde empirique. Lorsque Tal utilise le terme de modèle, il réfère à ce type de représentations abstraites des phénomènes qui sont utilisées à différentes fins scientifiques (prédire, expliquer, décrire, etc.; Tal 2012, voir p. 17-18 pour le détail). Dans le cadre de cette thèse, j'utilise le terme de la même façon.

Pour Tal, une philosophie de la mesure doit répondre à trois incontournables problèmes, offrant par ses réponses des considérations normatives à l'égard de la métrologie : le problème de la coordination, le problème de l'individuation des quantités (*comment peut-on savoir si deux procédures différentes mesurent la même quantité, notamment dans le contexte des incohérences entre les résultats qui émergent immanquablement?*) et le problème de l'incertitude (*comment l'incertitude est-elle gérée? Cette gestion est-elle épistémologiquement satisfaisante?*). Je focalise

mon attention sur le premier des trois problèmes, notamment parce qu'il est plus conséquent (les deux autres en émergent), mais aussi par souci d'espace et de temps.

L'approche centrée sur les modèles de Tal considère que la mesure d'une quantité physique consiste à imposer de la cohérence et de la régularité aux inférences faites à partir de processus physiques, *attribuant ainsi des valeurs à des paramètres au sein d'un modèle* (Tal 2012, 2016). Cette définition insistant sur le fait que la quantification issue de procédures d'observation établit une coordination entre phénomènes et paramètres d'un modèle plutôt qu'entre un phénomène et une valeur théorique associée à un standard (par exemple, le prototype international du kilogramme) contraste avec les approches traditionnelles. Cette distinction est importante puisqu'elle souligne que toute coordination entre une quantité théorique et une quantité physique, *même dans les cas pour lesquels des standards de mesure sont disponibles*, implique une charge théorique incarnée par les modèles utilisés. La définition, ou l'adoption par convention, d'une unité de mesure liée à un standard est insuffisante pour contourner cette situation.

Au premier regard, le tout semble tout à fait compatible avec la conception sémantique des théories scientifiques. Cela dit, les approches centrées sur les modèles se distinguent aussi de la conception sémantique des théories d'une manière importante pour la compréhension des conséquences du problème de la coordination. En effet, pour van Fraassen par exemple, la coordination est l'établissement de procédures qui mettent en relation un phénomène physique et de termes théoriques, éléments constitutifs de modèles. La relation cherchée est de nature isomorphe. Au contraire, Tal relâche ce critère :

Though not committed to any particular view on how models represent the world, the model-based account does not require models to mirror the structure of their target systems in order to be successful representational instruments. My framework therefore differs from the 'semantic' view, which takes models to be set-theoretical relational structures that are isomorphic to relations among objects in the target domain (Suppes 1960, van Fraassen 1980, 41-6). (Tal 2012, 18; voir aussi Mari 2005)

Ce relâchement par rapport à l'isomorphie modèle-phénomène s'explique par l'inextricabilité des modèles et des observations empiriques. Le modèle étant fondamentalement impliqué dans toute observation, comment serait-il possible d'établir une relation d'isomorphie

entre le modèle et ce qui est observé, comme si ces deux choses existaient indépendamment? Si une chose est observée, c'est qu'un modèle est impliqué de sorte à identifier quelles relations peuvent être exploitées pour quantifier un aspect d'un phénomène et à quelle notion théorique cet aspect correspond. Ce que les approches centrées sur les modèles mettent en exergue, c'est que tout cadre de référence dans lequel modèles et phénomènes sont pensés en isolement complet ou partiel est un cadre mal adapté à la description précise de la structure des pratiques scientifiques. Dans un contexte empirique comme celui des sciences occidentales, le modèle et les observations sont deux aspects d'une même chose.

Pour Tal, donc, aucune séparation forte n'existe entre observation et modélisation théorique, puisque la mesure *est* modélisation. Ce point marque une forte distinction entre les approches de Tal et de van Fraassen, qui se rejoignent pourtant sur l'impossibilité de comprendre l'aspect symbolique de la mesure sans passer par les modèles. Il faut aussi souligner que les approches centrées sur les modèles, comme l'approche sémantique de van Fraassen, considèrent que le critère permettant de distinguer les bonnes et moins bonnes procédures métrologiques est l'adéquation avec l'expérience. Chez Tal, cependant, cette adéquation s'établit autrement que par isomorphie.

Dans l'approche de Tal, l'adéquation empirique d'une procédure relève de sa capacité à établir l'objectivité de l'observation, c'est-à-dire la possibilité de dire qu'elle nous informe à propos de l'*objet* mesuré. Pour Tal, ce but épistémique est atteint lorsqu'un modèle se cristallise de sorte à standardiser les procédures de mesure, ce qui inclut la correction des imprécisions caractéristiques des situations métrologiques individuelles :

Prior to their representation by an idealized model, there is no way of testing whether different instruments measure the same quantity; any agreement or disagreement among their indications may be construed as coincidental and attributed to some local feature of the instruments or environments. It is only once their idiosyncrasies are idealized away in a mutually coherent fashion that instruments can be viewed as sources of objective knowledge about a common quantity, such as temperature or frequency. (Tal 2017, 14)

La modélisation permet ainsi de tirer des conclusions générales, en donnant aux descriptions un sens qui dépasse l'idiosyncrasie des pratiques isolées. Des chercheurs et chercheuses à différents endroits dans l'univers peuvent ainsi être confiants et confiantes que

leurs mesures veulent dire la même chose malgré les spécificités de chaque situation. Dire de deux choses qu'elles sont de la même longueur acquiert une valeur objective grâce aux modèles et à l'abstraction qu'ils impliquent et qui corrigent l'influence des circonstances individuelles (liées aux instruments ou au contexte).

Pour bien conceptualiser la chose de concert avec Tal, il faut insister sur la double nature du phénomène étudié. Toute procédure de mesure est constituée d'éléments physiques en relation les uns avec les autres et d'aspects symboliques qui composent les modèles représentant lesdits éléments physiques (les deux aspects à coordonner qui sont, pour van Fraassen, potentiellement isomorphes). Les éléments physiques incluent des interactions entre des instruments de mesure, des échantillons à mesurer, l'environnement et des êtres humains. Les aspects symboliques incluent un ensemble d'opérations inférentielles qui sont structurées sous forme de modèles, dont l'étude permet d'extraire les prémisses (*assumptions*) qui guident l'observation empirique.

Une autre dichotomie doit être considérée, soit celle impliquant l'indication d'un instrument de mesure et le résultat d'une procédure de mesure. L'indication relève entièrement de l'aspect physique d'un processus de mesure. Il s'agit par exemple d'une colonne de mercure dans un cylindre de verre, ou du bras d'un ampèremètre. Le résultat d'un processus de mesure, au contraire, est issu du traitement de ces indications sur la base des modèles qui structurent les inférences. Par exemple, la hauteur d'une colonne de mercure peut être utilisée pour inférer la température de la substance dans laquelle elle est plongée. Il est important de noter que ces inférences sont souvent occultées (*black boxed*) par des instruments utilisés couramment. C'est le cas lorsqu'un thermomètre mobilise une colonne de mercure dont la hauteur peut être directement interprétée en termes de degré Celsius grâce à la gradation du cylindre de verre dans lequel elle se trouve. Il s'agit bien alors d'un thermomètre au mercure qui a été confectionné en acceptant de fait le modèle au sein duquel pression, volume et température sont des paramètres corrélés, c'est-à-dire des quantités théoriques en interaction.

Ces deux dichotomies permettent de se représenter très clairement les difficultés reliées au problème de la coordination : toute procédure observationnelle impliquant la mesure relève

d'instruments dont les indications ne peuvent être coordonnées à des énoncés théoriques que par l'entremise d'une modélisation préalable; les modèles théoriques impliqués sont corroborés par les observations empiriques; cette circularité superficielle repose en fait sur les itérations passées de l'interaction entre aspects physiques et symboliques des procédures de mesure. Je suggère donc de penser l'approche de Tal comme relevant d'un cohérentisme progressif (Chang 2007) dont les itérations épistémiques traduisent des pratiques de modélisation dont les transformations successives sont déterminées notamment par le rôle que jouent les phénomènes dans l'appareillage physico-théorique en question. L'impératif du progrès mène ainsi les praticiens et praticiennes à peaufiner les modèles de sorte à mieux répondre aux problèmes qui ont émergé des itérations passées relatives à ces modèles.

Tal et les autres tenants et tenantes des approches centrées sur les modèles sont tout à fait conscients et conscientes qu'il existe une grande variété de déterminants qui rendent possibles les procédures métrologiques et qui les font progresser au travers du temps. Par exemple, des avancées techniques, peu importe comment elles sont réalisées, permettent et imposent éventuellement le développement de nouveaux modèles (au chapitre 7, par exemple, je présente l'utilisation de modèles réticulés en phylogénétique comme étant *notamment* le fruit des puissances calculatoires développées dans les dernières décennies). La prise en compte de nouveaux phénomènes peut aussi grandement influencer les pratiques de mesure (par exemple, l'accent mis sur les transferts génétiques entre lignées a aussi contribué à populariser les modèles réticulés en phylogénétique). Ainsi, il ne s'agit aucunement d'affirmer l'indépendance des modèles face à l'empirie. Il est plutôt question de souligner la nécessité de la modélisation pour appréhender le domaine empirique ainsi que sa capacité à guider les progrès scientifiques, tout en insistant sur les conséquences du problème de la coordination : les modèles sont sous-déterminés par l'empirie de sorte que l'épistémologie se doit de considérer d'autres facteurs pour expliquer et décrire le dynamisme scientifique.

Ainsi, les points saillants de l'approche de Tal sont les suivants : 1) la mesure consiste à attribuer une valeur à un paramètre au sein d'un modèle; 2) cette attribution se fait par l'entremise de procédures détenant un aspect physique et un aspect symbolique; 3) en modélisant les processus physiques impliqués, les pratiques métrologiques font abstraction des

spécificités locales afin de créer des observations objectives qui permettent d'associer de manière standardisée des quantités physiques et des concepts théoriques. Les modèles sont ainsi au cœur des processus de standardisation de la pratique : l'observation (scientifique, précise, etc.) est modélisation.

2.3.4 Une critique des approches cohérentistes

Dans son article de 2019, Isaac propose une approche réaliste de la mesure présentée en deux temps. Il formule d'abord une critique des approches cohérentistes des trois auteurs dont les idées ont été résumées précédemment dans ce chapitre, puis il se base sur la notion de succès métrologique pour structurer une posture réaliste concernant l'existence de points fixes comme la température d'ébullition de l'eau. Isaac accepte l'inévitabilité du problème de la coordination, auquel il réfère par la notion de boucle épistémique. Selon lui, ces boucles épistémiques sont en revanche tout à fait compatibles avec une approche réaliste.

Isaac est loin d'être le premier à défendre une approche réaliste en philosophie de la mesure. Traditionnellement, le réalisme en philosophie de la mesure s'est révélé être une posture extrêmement forte, allant plus loin que le réalisme scientifique que l'on rencontre habituellement en philosophie des sciences. En effet, selon ce courant de la philosophie de la mesure, non seulement les théories scientifiques traduisent adéquatement la structure réelle sous-jacente aux phénomènes, mais même la teneur quantitative de leurs concepts serait réelle (voir notamment Byerly et Lazara 1973, Mundy 1987 ou Swoyer 1987). Par exemple, la température serait un concept qui renvoie effectivement à un aspect de la réalité et, de surcroît, cet aspect de la réalité serait bel et bien une quantité, c'est-à-dire qu'une échelle quantitative adéquate serait apte à le décrire *parce que* cette échelle capture la nature du phénomène tel qu'il existerait indépendamment de nos gestes épistémiques. En cohérence avec cette croyance à la base de leurs travaux, les auteurs et autrices associé-e-s à ce courant tendent plutôt à réfléchir au statut métaphysique des concepts quantitatifs et à la structure mathématique des échelles métrologiques qu'à interroger la structure justificative des sciences mobilisant des observations empiriques (Isaac 2019; Tal 2012, 2015). Leurs travaux suggèrent une adhérence *de facto* à la

croyance selon laquelle la mesure (et autres observations empiriques) fournit un accès direct à la réalité et offre du même coup une fondation absolue à la connaissance.

Dès le début du texte, Isaac prend soin de se distancier de ces approches naïves :

The view defended here reverses this priority, articulating a realism about measurement outcomes weaker than traditional realism. In particular, I argue that the convergent assignment of increasingly precise values that constitutes successful measurement serves as incontrovertible evidence for *fixed point* in the world [...] even though the evidence it provides for any nonnumerical theoretical description of these points is defeasible. (Isaac 2019, 931)

L'auteur défend un réalisme limité à l'existence de points fixes *réels* « dans le monde », comme la température d'ébullition de l'eau pourrait l'être. Il précise que les théories qui étoffent l'aspect mathématique des opérations métrologiques sont toujours réfutables, mais que les points fixes sont quant à eux des réalités que les entreprises métrologiques peuvent identifier en dépassant la charge théorique des observations : « I call it realism because I take fixed points to exist independent of our practice and metric representation of them to outstrip “observable” phenomena. » (Isaac 2019, 940).

Selon Isaac, les approches cohérentistes de Chang, de Tal et de van Fraassen nient le réalisme et la capacité de la mesure à offrir des résultats objectifs. Pour lui, l'objectivité est synonyme de l'indépendance des démarches scientifiques par rapport à des intérêts qui pourraient être décrits comme étant subjectifs, propres aux chercheurs et chercheuses ou à leur communauté dans son ensemble. D'après cette définition, les trois auteurs cohérentistes rejettent en effet l'objectivité. Ce commentaire doit toutefois être fortement nuancé puisqu'Isaac offre, sur cette base, une lecture défavorable des auteurs qu'il critique (et qu'il érige ainsi en épouvantails). Par exemple, lorsqu'il transmet sa compréhension des travaux de Tal et de Chang, il affirme que ceux-ci priorisent strictement la cohérence interne des modèles (et théories) plutôt que leur confrontation à l'expérience afin d'établir leur validité : « The lessons of this example are analogous to those of epistemic iteration: measurement improvements appears to rest on internal standards of coherence rather than on correspondence with external quantities. » (Isaac 2019, 934). Pourtant, Tal et Chang décrivent très explicitement la rencontre entre, d'un côté, les modèles (chez Tal) et théories changeantes (chez Chang), et de l'autre, l'expérience.

D'une part, Tal présente cinq principes d'ordre général qui doivent selon lui guider les analyses philosophiques de la mesure. Le quatrième principe se lit comme suit : « Success in stabilizing measurement standards cannot be explained by reduction to either natural or socio-historical factors; instead, stability results from the iterative entanglement of these factors in the process of standardization. » (Tal 2016, 319). Plus généralement, sa description de la coordination du Temps Universel Coordonné (*Universal Time Coordinated*) et de la standardisation de la seconde comme unité de mesure temporelle révèle une panoplie de rencontres entre expérience et modèles (par exemple, la gestion des horloges atomiques dont les périodes dérivent progressivement de leur calibration initiale ou encore la mise en commun, par la modélisation, des multiples horloges utilisées pour coordonner le temps à l'échelle planétaire) qui sont cruciales pour la pratique métrologique qu'il dissèque. L'article de 2016 est son texte le plus complet à ce sujet, mais n'est malheureusement *pas* cité par Isaac.

D'autre part, Chang rappelle sans cesse, dans son livre *Inventing Temperature* de 2004, que l'expérience et le monde concret qui la sous-tend sont la matière de tout labour scientifique, même si la philosophie et l'histoire des sciences gagnent à décrire ces pratiques sur la base d'une approche cohérentiste. En effet, Chang, en discutant des points thermométriques fixes, ne pourrait être plus tranchant à ce sujet : « Fixed points can be artificially created in the same way seedless watermelons can be created; these things cannot be made if nature will not allow them, but nonetheless they are our creations. » (Chang 2004, 49).

Van Fraassen n'est pas critiqué de front comme les deux autres auteurs, puisqu'aucune section de l'article ne lui est réservée. La citation suivante est toutefois révélatrice de l'avis qu'Isaac se fait de son approche : « Consequently, pace van Fraassen, attempts to increase precision are empirically constrained, since this directed metric for improvement can only be satisfied through the cooperation of nature » (Isaac 2019, 937). J'invite le lecteur ou la lectrice à comparer cette charge, selon laquelle van Fraassen considère que le progrès de la précision métrologique est libre face à l'empirie, avec la citation suivante qui démontre bien que Isaac fait une interprétation douteuse du cohérentisme :

To be sure, this requirement of coherence is not simply one for logical consistency.
Whether a sort of mechanism can be used to define the family of standard clocks

depends on empirical regularities that may or may not obtain. The central coherence condition on the family of standard clocks, recall, was that if two are in coincidence, they run at the same rate, they run in synchrony. That is a matter of empirical fact. But if two sorts of mechanism satisfy this condition, there is no matter of fact as to which runs evenly, and a choice or convention alone can decide on one of them. (van Fraassen 2008, 136)

Ce retour rapide sur les approches cohérentistes démontre que la critique faite par Isaac à leur égard est construite sur un constat erroné. Le texte érige épouvantails avec les écrits de Chang, de Tal et de van Fraassen, critiquant une priorisation exagérée de la cohérence interne qui est pourtant inexistante. Le cohérentisme, comme défini par les auteurs étudiés, est une posture philosophique et historique qui nie que l'expérience puisse faire office de base ou de justification *absolue* à la connaissance puisque le rapport à l'expérience est influencé par le contexte dans lequel les agents et agentes épistémiques opèrent. Une approche cohérentiste qui considère que l'expérience est un déterminant crucial des dynamiques scientifiques est cependant tout à fait possible et avérée.

Il convient maintenant d'analyser la défense du réalisme, pour laquelle Isaac mobilise deux arguments gravitant autour de la notion de succès métrologique (*successful measurement*). Premièrement, la convergence de différentes méthodes sur certains points fixes suggère que ces points fixes existent indépendamment des pratiques épistémiques puisqu'elles mobilisent des procédures et des prémisses implicites ou explicites complètement différentes. Deuxièmement, le fait qu'une méthode soit précise, c'est-à-dire qu'elle renvoie à des résultats cohérents à chacune de ses utilisations, demeure inexpliqué en l'absence de valeurs réelles qui influent sur les résultats de mesure.

L'idée selon laquelle un réalisme pourrait être défendu sur la base de ces deux constats est inexacte. Cette inexactitude survient parce qu'Isaac appréhende les dynamiques métrologiques de manière anhistorique, comme les philosophes des sciences conventionnalistes le faisaient il y a cent ans. Les pratiques scientifiques sont pourtant des objets éminemment historiques; une démarche qui nie cette caractéristique perd donc nécessairement de l'information importante. Dans ce qui suit, j'abonde en ce sens, en répondant aux deux arguments principaux de l'article d'Isaac.

Considérons tout d'abord l'argument de la convergence. Prenons un exemple étudié en détail par Chang dans son livre de 2004 auquel Isaac réfère dans son article. Cet exemple relève de la « découverte » du point fixe que représente la température d'ébullition de l'eau (18^e siècle). Avant tout, il faut se rappeler que l'établissement de points fixes est nécessaire pour le calibrage de thermomètres, pour la construction d'échelles et donc pour l'étude de la chaleur et de la température. Nul besoin de détailler ici les va-et-vient complexes entre théories, observations et expérimentations qui ont permis aux scientifiques de l'époque de converger vers le point fixe que l'on accepte aujourd'hui volontiers comme étant de 100 °C. Le premier chapitre de *Inventing Temperature* (Chang 2004) détaille le tout avec bien plus de précision que je ne saurais le faire. Il suffit de souligner quelques caractéristiques du problème pour comprendre en quoi ce développement scientifique ne peut que superficiellement être décrit comme une découverte soutenant la thèse selon laquelle les points fixes sont athéoriques et indépendants de nos pratiques.

Lorsqu'une quantité d'eau est en ébullition, sa température est hétérogène; elle peut présenter des écarts de plus de 10 °C en fonction de l'endroit où un thermomètre sonde le liquide. L'eau qui bout légèrement et celle qui s'agite à grands bouillons ont aussi des températures différentes : l'eau bout en effet à des températures différentes en fonction de la pression qui est exercée sur elle. En ce qui concerne la vapeur fraîchement produite, comme sa contrepartie liquide, elle a une température variable. Bien que mesurer sa température dès que possible après l'évaporation ait fortement contribué à l'identification du point fixe, il n'en demeure pas moins qu'il a fallu faire abstraction d'une grande variation face aux idiosyncrasies caractérisant chaque mesure pour « découvrir » un point métrique fixe comme celui du point d'ébullition. Toute cette problématique est bien sûr doublée du fait que deux thermomètres, qu'ils utilisent la même substance thermométrique ou non, finissent inlassablement par offrir des résultats divergents que les modèles doivent niveler pour créer des mesures « objectives ».

En bref, la « découverte » du fameux 100 °C nécessite la mise en place de nombreuses conventions par rapport aux conditions matérielles qui rendent possible la stabilité des phénomènes, au traitement des résultats divergents et aux aprioris théoriques mobilisés :

If defensible fixed points do not occur naturally, they must be manufactured. I do not, of course, mean that we should simply pretend that certain points are fixed when they are not. What we need to do is find, or create, clearly identifiable material circumstances under which the fixity does hold. (Chang 2004, 49)

Il faut créer la fixité, de la même manière que l'on crée des melons sans pépins.

Dès que l'on traite de la question en tenant compte de sa dimension historique, on se rend bien compte que la découverte du point d'ébullition est un accomplissement louable qui a mis plus de 100 ans à se mettre en place, du moins au sein de la communauté scientifique occidentale. Cet accomplissement implique une part indéniable de théorie, c'est-à-dire que divers modèles ont été mobilisés pour gérer les divergences entre les résultats, pour défendre l'utilisation d'un point fixe relié à l'eau plutôt qu'au beurre ou à la cire par exemple, pour mieux circonscrire l'environnement adéquat pour mener des expériences, pour peaufiner l'art de la création d'outils thermométriques, etc. (Chang 2004). Grâce à de nombreux débats, un phénomène diffus a été transformé en connaissance précise; un point fixe a été créé. Le processus historique qui rend le tout possible *précède* la « découverte » du point fixe, qui ne peut donc être considéré comme étant indépendant de ce processus qu'en faisant fi du phénomène étudié, soit la structure justificative des sciences.

L'erreur d'Isaac est donc bien de sortir ce genre d'accomplissement, soit l'établissement d'un point de référence métrologique, de son contexte historique. Nonobstant cette faille, suivons tout de même son raisonnement un peu plus loin. D'un point de vue anhistorique, il semble en effet, au premier regard, que la présence d'un point fixe réel permettrait d'expliquer pourquoi différentes méthodes convergent : elles convergent sur le point fixe, une réalité qui n'attend que d'être découverte. Cette logique est tout à fait cohérente, mais elle n'acquiert une validité quelconque que si et seulement si de tels points fixes existent effectivement indépendamment des méthodes qui convergent sur eux. La validité du raisonnement repose donc ultimement sur la possibilité de démontrer que ces points fixes existent indépendamment de notre pratique. La question qui se pose alors est la suivante : est-ce que la convergence de deux ou plusieurs méthodes sur un point fixe est une preuve suffisante de l'existence indépendante de ce point fixe? Il me semble que poser la question, c'est y répondre, mais reprenons tout de même l'exercice sous la direction de Tal.

L'argument d'Isaac omet un élément important des dynamiques en jeu, soit l'idée que la convergence, tout comme l'établissement de points fixes, est le fruit d'un travail soutenu et effectué en amont par la communauté scientifique. Ce travail implique notamment des développements théoriques. Le passage suivant illustre ce point avec éloquence :

Meaningful agreement or disagreement can be established among measurement outcomes because outcomes are already corrected for known distortions and associated with uncertainty estimates for unknown distortions. By contrast, instrument indications are not directly comparable to each other for two reasons. First, different instruments often display their outputs differently, for example, by the position of a needle relative to a dial, occurrence of digits on a display, the frequency of a sound, and so on. Second, even when indications are converted to numbers, their convergence is neither necessary nor sufficient for agreement among measurement outcomes. *After applying required error corrections*, instruments that generate the same indications may be interpreted as producing incompatible measurement outcomes and vice versa. For example, when primary frequency standards are compared to each other, each of their raw output frequencies is first corrected for a host of systematic errors. As different clocks differ in their environment and design, they are assigned different corrections factors. Consequently, two primary standards that 'tick' at the same rate would normally be considered to be in significant disagreement. It is only their already-adjusted measurement outcomes that are considered meaningful for comparison. (Tal 2016, 320, italiques ajoutés)

Je prends sur ce point position du côté des auteurs et autrices cohérentistes et j'exhorte le lecteur ou la lectrice à en faire de même : la convergence de différentes méthodes est une preuve insuffisante pour défendre une posture réaliste par rapport à l'existence de points fixes puisque toute croyance selon laquelle ces points fixes existent dépend des méthodes convergentes (sinon, elle est non fondée). Autrement dit, soutenir la thèse réaliste à l'égard des points fixes requiert une coordination entre phénomènes et théorie, une coordination qui s'inscrit au sein de boucles épistémiques. Même si deux procédures réalisant la coordination convergent sur certains résultats, cette convergence est, de fait, teintée par ces boucles épistémiques. Les résultats ainsi acquis demeurent dépendants des mécanismes qui permettent de surmonter les incarnations du problème de la coordination.

Il est intéressant de noter que la situation décrite par Tal s'applique, *mutatis mutandis*, à la phylogénétique. Par exemple, la convergence de méthodes d'analyse phylogénétiques aussi distinctes que celles basées sur des données moléculaires et celles basées sur des données morphologiques sur des résultats communs est le fruit de labeurs *visant explicitement à établir*

une telle convergence. Il serait donc tout à fait malvenu de considérer cette convergence, objectif assumé des praticiens et praticiennes, comme reflétant un fait existant indépendamment des praticiens et praticiennes : « Discordance between trees drives the field forward as all of its practitioners attempt to find consensus between and explanations of conflicting phylogenies. » (O'Malley 2016, 78). Les points fixes expliquent la convergence, non pas en tant que réalités ontologiques, mais simplement en tant que motivations, en tant qu'objectifs épistémiques.

Il me semble donc plus judicieux d'adopter, autant dans le contexte d'une épistémologie de la mesure que dans celui d'une épistémologie de la classification, un modèle épistémique pour appréhender les modalités des boucles épistémiques et leurs transformations au travers du temps que de nier ou de minimiser leur importance par une approche réaliste. Le réalisme suggère en effet que la justification ultime de la validité des procédures scientifique est « naturelle » ou indépendante de ces procédures, alors que les phénomènes, accessibles seulement par l'entremise de ces procédures, ne forment qu'une contrainte déterminant les itérations épistémiques parmi tant d'autres. Dans une telle perspective, l'argument de la convergence d'Isaac est inapte à démontrer la réalité des points fixes et encore plus inapte à démontrer l'utilité d'une posture réaliste pour l'analyse épistémologique de pratiques scientifiques.

Le deuxième argument d'Isaac mobilise la notion de précision. Selon lui, le fait que certaines pratiques soient plus précises que d'autres ou que nous puissions améliorer la précision d'une procédure requiert la réalité indépendante de points fixes. En effet, nous dit Isaac, la précision implique de se rapprocher d'une cible. Que pourrait être cette cible sinon un point fixe réel? Cette logique est encore une fois incompatible avec la pratique parce que le problème de la coordination existe et parce qu'il faut conséquemment avoir des prémisses théoriques pour établir des cibles à partir desquelles il est possible d'évaluer la précision d'une procédure.

Cela dit, les problèmes de l'argument d'Isaac vont plus loin. Selon lui, la notion de précision joue son rôle par l'entremise des représentations géométriques des résultats de mesure. En effet, il considère que les pratiques de mesure assignent des valeurs quantitatives aux phénomènes, puis représentent ces mesures sur un espace géométrique. Selon Isaac, *cet espace a une structure*

qui serait antécédente à toute observation. Cette structure fixe la relation entre deux points : un intervalle d'une valeur donnée, dans cet espace métrique, sera toujours égal à un autre intervalle de même valeur. Cette structure contraint les résultats de la mesure puisqu'elle procure un standard mathématique et non physique, un ensemble de contraintes formelles auxquelles doivent se conformer les pratiques métrologiques :

This antecedent structure constraints the relationship between measurement outcomes, independently restricting our assessment of them as the same or different, or converging or not, in a manner impervious to ad hoc revision. [...] Thus, the metric of this space serves two functions: (i) it represents the distances between different measured quantities, but (ii) it also provides a directed metric for improving measurement of a single quantity. (Isaac 2019, 937)

Face à cette citation, il est légitime de se poser la question suivante : « Cet espace métrique structuré ne serait-il pas un a priori théorique (un ensemble de prémisses structurées guidant les opérations subséquentes) que les métrologues adoptent et qui incarne bien la teneur du problème de la coordination? » Van Fraassen répond positivement à cette question. En effet, alors qu'il traite de la possibilité de répondre au problème de la coordination en présumant que le monde peut être décrit géométriquement avant toute coordination, une solution qui correspond à celle d'Isaac, van Fraassen nous dit la chose suivante :

Reichenbach would be quite right to retort that this parlance is meaningful only *after* the problem of coordination is solved. It cannot be a solution! For this 'solution' begins by thinking of the WORLD as a structure, a set of elements with certain relations between those elements, described precisely in the terms we use in geometry. So it takes for granted that we can represent the WORLD mathematically. And what is that, if not to say that there is a certain 'representing' relationship between the WORLD and some mathematical structure? – thus assuming precisely what was to be illuminated. (van Fraassen, 137-138)

La « structure antécédente » d'Isaac ne peut contraindre la mesure que si une coordination est déjà en place et permet de faire le passage du plan mathématique aux phénomènes. Si cette structure est *effectivement* antécédente aux boucles épistémiques, elle n'a alors aucun contenu empirique et est incapable de guider la métrologie et de contraindre ses résultats. Conséquemment, elle peut difficilement établir la validité d'une posture réaliste par rapport à des points fixes au sein de l'empirie qui dépasserait ces boucles épistémiques.

Chang, Tal, van Fraassen ainsi que les auteurs et autrices classiques amorcent leurs réflexions métrologiques en appréhendant une difficulté de base, inhérente à toute épistémologie de la mesure. Il s'agit de rendre compte du passage entre des phénomènes physiques et un monde symbolique dans lequel les contraintes formelles sont faciles à imaginer, à expliciter et à justifier. Pour Isaac, ce passage va de soi; la réalité des points fixes dialogue sans problème avec la structure formelle de l'espace métrique. Or, l'histoire des sciences révèle incontestablement que l'acceptation apriorique de principes ontologiques, de prémisses implicites ou explicites, d'engagements, d'énoncés théoriques, de théories, de modèles (ou appelez-les comme vous le souhaitez) est nécessaire pour l'observation quantitative (et toute autre forme d'observation, mais restons concentré·e·s sur la métrologie pour l'instant). C'est seulement à la lumière de telles croyances que l'on peut associer un état des choses et un espace métrique.

L'erreur fondamentale et impardonnable d'Isaac (d'un point de vue philosophique) est d'avoir contourné le problème de la coordination et d'avoir défendu une posture incompatible avec celui-ci. Soyons clairs : le problème de la coordination n'implique *pas* pour autant une posture antiréaliste; cependant, il force très certainement à reconnaître que le monde ne nous est accessible que par le prisme de nos modèles, théories, énoncés et croyances qui rendent possible l'observation scientifique. C'est donc la question même du réalisme qui perd son sens : « Consider what effects, which might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of those effects is the whole of our conception of the object. » (Peirce 1992, 132). Que des objets comme les points fixes soient réels ou non, nous continuerons de les observer, les établir et les décrire d'une manière tout à fait dépendante de nos modèles, théories, pratiques, etc. Leur réalité (ou son absence) est ainsi dénuée d'effets concrets (« *practical bearings* »), de sorte que le sens même de cette réalité s'évapore face à l'inévitabilité de la charge théorique des observations. C'est probablement ce qui explique pourquoi, à l'exception d'Isaac, la philosophie de la mesure contemporaine vise à décrire les pratiques scientifiques plutôt qu'à postuler l'existence indépendante de choses comme les points

fixes, une indépendance dont la démonstration, par définition, ne peut que rester épistémiquement hors d'atteinte¹⁵.

2.4 Conclusion

Alors qu'au chapitre précédent j'ai démontré que le problème de la coordination s'incarne bien au cœur de la toile épistémique qu'est la phylogénétique, je me suis efforcé, dans ce chapitre-ci, d'explorer les conséquences de ce problème que les philosophes des sciences et plus précisément les philosophes de la mesure ont analysé depuis un siècle. J'y présente d'abord les approches dites traditionnelles (conventionnalisme, opérationnalisme et constructivisme), principalement pour que le lecteur ou la lectrice puisse comprendre de quelle manière les postures contemporaines sur le sujet ont été sculptées. Le conventionnalisme, une approche fondationnaliste, considère que le problème de la coordination engendre une circularité justificative qui mine les aspirations empiristes des pratiques scientifiques. Or, des auteurs comme Mach, Poincaré et Reichenbach considèrent que cette circularité est finalement un épiphénomène justificatif : la solidité de l'édifice scientifique repose sur des conventions, elles-mêmes extérieures à la connaissance. Pour Reichenbach, ces conventions sont des principes définitionnels, des éléments sémantiques sur lesquels la connaissance peut s'appuyer sans souci. La connaissance empirique, malgré le problème de la coordination, est ainsi fondée de manière absolue.

Au contraire, le constructivisme, du moins lorsqu'il est lu naïvement comme je me suis engagé à le faire à des fins heuristiques, considère qu'il est impossible de fonder les connaissances

¹⁵ Il m'a été suggéré à quelques reprises qu'il est possible d'adopter un réalisme minimal en toute cohérence avec le problème de la coordination et ses conséquences. Ce réalisme minimal consiste à postuler une réalité existante indépendamment de l'esprit humain tout en acceptant son inaccessibilité épistémique. Il faudrait donc demeurer agnostique par rapport à la relation entre théories scientifiques et cette réalité indépendante. Je conçois bien la cohérence entre cette option et l'implacabilité du problème de la coordination. J'ai plus de difficulté à concevoir comment une telle croyance, infondée de fait (puisque'il est accepté qu'on n'a aucun accès indépendant à cette réalité), nous aide à réaliser des analyses épistémologiques efficaces. De plus, cette posture réaliste tend à amalgamer deux phénomènes distincts : celui selon lequel philosophes et scientifiques *croient* qu'il y a une réalité indépendante à laquelle peuvent correspondre leurs théories et modèles et celui selon lequel une telle réalité existe ou non. Pour ces deux raisons, je conclus que le réalisme (et l'antiréalisme) ne mérite aucune place au sein du modèle que je préconise pour l'analyse épistémologique.

empiriques de manière absolue. Les conséquences du problème de la coordination sont telles qu'il ne reste plus, d'un point de vue épistémologique, qu'à comprendre comment les dynamiques sociohistoriques caractérisant les sciences en viennent à déterminer quels énoncés, hypothèses, théories ou pratiques peuvent être considérés comme étant valides.

En contraste avec ces deux approches qui dialoguent avec l'inévitabilité (superficielle ou non) du problème de la coordination, l'opérationnalisme de Bridgman, tout comme les postures réalistes en philosophie de la mesure, nie la pertinence de ce casse-tête épistémique et donc, du même coup, la pertinence de la charge théorique des observations. Pour ce physicien, les opérations concrètes dictent complètement la teneur de la dimension théorique des sciences, chaque concept théorique ne détenant que le sens qu'on peut lui attribuer sur la base de l'ensemble des procédures auxquelles il est associé. Cette approche néglige cependant un fait important : les procédures en question sont elles-mêmes chargées théoriquement, de sorte qu'il est inexact de parler d'une influence unidirectionnelle reliant procédures d'observation et théorie.

L'étude de ces trois courants de pensée permet d'identifier des desiderata pour une épistémologie qui s'intéresse aux pratiques scientifiques. À l'encontre du conventionnalisme, elle doit reconnaître l'historicité des pratiques scientifiques. Cette historicité permet de comprendre comment les connaissances sont construites malgré la présence du problème de la coordination et assure une description des dynamiques scientifiques faisant preuve de plus d'exhaustivité. Ensuite, et contrairement au constructivisme naïf, elle doit être en mesure de rendre compte de la normativité en sciences, et du même coup, du progrès des sciences ainsi que de son ancrage empirique. C'est finalement l'espoir de conjuguer ces deux desiderata qui a mené à l'élaboration de nouvelles approches en philosophie de la mesure au tournant du 21^e siècle.

Après avoir établi ces desiderata, j'ai exploré trois réponses contemporaines au problème de la coordination, soit celles de Chang, de van Fraassen et de Tal. Ces trois approches ont en commun non seulement la reconnaissance du problème en question, mais aussi la croyance selon laquelle les structures justificatives en sciences sont des tous cohérents dont les dynamiques s'érigent à la rencontre d'une variété de facteurs. Au nombre de ces facteurs, on retrouve des

éléments sociohistoriques ainsi que la confrontation entre les théories et l'empirie. Ces approches, qui sont donc éminemment empiristes, nient cependant la possibilité d'établir des fondements absolus à la connaissance, en cohérence avec le va-et-vient entre phénomènes et théorie identifié par l'entremise du problème de la coordination. Le chapitre trouve son terme dans l'exploration d'une critique de ces approches cohérentistes (auxquelles je suggère de référer en bloc par l'appellation « cohérentisme historique »), une critique dont j'illustre les points faibles : d'une part, une mauvaise lecture des auteurs ciblés (Chang, van Fraassen et Tal), et d'autre part, la croyance selon laquelle des points fixes pourraient expliquer la convergence et la précision de méthodes d'observation, alors même qu'ils sont les produits de ces méthodes.

Dans les deux prochains chapitres, je couvre ce qui est, à ce point-ci de l'ouvrage, un angle mort de mon cadre théorique pour l'analyse des pratiques phylogénétiques. J'y appréhende la riche notion d'espèce naturelle souvent mobilisée pour traiter de l'homologie et des espèces biologiques (ainsi que d'autres unités taxonomiques). Plus spécifiquement, je défends dans ces chapitres le rejet de la notion d'espèce naturelle sur la base de son incapacité à guider l'analyse des pratiques classificatoires en général et de la phylogénétique en particulier. Il convient avant tout de résumer les outils conceptuels mis de l'avant par Chang, van Fraassen et Tal qui sont quant à eux retenus pour l'analyse de la phylogénétique. Ces outils sont aussi au cœur de l'analyse du premier chapitre de cette thèse bien que leur utilisation y soit implicite.

Les outils issus du cohérentisme progressif de Chang :

- La notion d'itération épistémique : il s'agit de penser le dynamisme scientifique comme étant constitué d'une succession d'itérations épistémiques, au sein de laquelle chaque état de connaissance puise dans les ressources fournies par le précédent.
- Les états de connaissance : il s'agit des itérations qui forment la matière première étudiée par l'épistémologie. Le terme est utilisé vaguement par Chang de manière à permettre l'identification pragmatique de discours scientifiques dont l'étude pourrait se révéler informative. Un état de connaissance peut être englobant (par exemple, la phylogénétique qui mobilise les modèles arborescents; un tel état de connaissance peut être mis à jour par l'étude d'ouvrages de référence, par une revue du corpus

disciplinaire, etc.) ou plus limité (par exemple, une reconstruction phylogénétique spécifique utilisant le modèle arborescent communiquée dans un article).

- L'impératif de progrès : le moteur du changement est l'injonction à améliorer les itérations successives. Cette amélioration peut être évaluée par l'entremise d'une pluralité de critères ou de valeurs épistémiques à spécifier en amont. L'ensemble de critères ou de valeurs est sujet à changement au fil des itérations.
- Le principe de respect : malgré l'injonction à l'amélioration des systèmes, les praticiens et praticiennes tendent à « respecter » les itérations passées qu'ils et elles cherchent à dépasser, faisant ainsi preuve d'un conservatisme qui contribue à la légitimité épistémique des pratiques.
- Les principes ontologiques : il s'agit de croyances comparables aux gonds de Wittgenstein en ce sens qu'elles forment des points de pivot irremplaçables pour que la structure épistémique d'un champ gnoséologique permette les inférences auxquelles il aspire. Ces croyances sont injustifiées et parfois injustifiables; elles sont donc acceptées *a priori*.

Les apports de van Fraassen :

- La vision surplombante : cette vision incite l'épistémologue à faire abstraction de l'historicité de son objet d'étude afin de lire des pratiques scientifiques comme une théorie scientifique stable le suggère. Il s'agit ainsi de mettre entre parenthèses le problème de la coordination afin de cerner les relations entre les éléments d'une structure épistémique d'après la logique propre à la théorie scientifique stabilisée en question.
- La vision interne : cette vision consiste à étudier les pratiques scientifiques en priorisant leur historicité et se rapproche de l'approche de Chang. Elle consiste à lire les relations entre les éléments constitutifs d'une structure épistémique au-delà de la logique d'une éventuelle théorie stabilisée, considérant de fait que certaines associations d'idées peuvent être le produit de déterminants sociohistoriques tout autant que le fruit d'une saine rencontre entre logique, théorie et empirie.

- La vision synoptique : van Fraassen exhorte finalement ses lecteurs et lectrices à mettre en commun ces deux regards sur les pratiques scientifiques afin d'offrir la lecture la plus complète possible du système étudié.

Les approches centrées sur les modèles :

Ces approches soulignent le rôle central des modèles qui permettent aux observations scientifiques d'être effectuées de manière standardisée, de sorte qu'elles acquièrent ainsi une certaine dose d'objectivité. Ces approches offrent une manière précise de mobiliser le cohérentisme de Chang en étudiant les itérations successives des modèles qui rendent un système épistémique effectif. De surcroît, ces approches soulignent que toute observation scientifique implique un travail de modélisation, de sorte qu'il est futile de conceptualiser les observations ou les objets observés indépendamment des modèles qui les rendent possibles ou observables.

Chapitre 3 – L’arche des espèces naturelles

“It’s another good reason to speak to Inaya. Maybe you can figure her out.”
Nyx laughed. “Because I’m a woman?”
“What?” He looked genuinely affronted.
“Khos, me and Inaya have nothing in common. When are you going to get that?”
- Kameron Hurley, dans *Infidel*

3.1 Introduction

Les théories des espèces naturelles constituent le principal arsenal philosophique disponible pour appréhender les pratiques classificatoires (Brigandt 2020; Conix et Chi 2020; Kendig 2016). L’idée générale qui émane de ce corpus de textes est que la notion d’espèce naturelle peut nous aider à comprendre les succès épistémiques des pratiques classificatoires par la description de la teneur métaphysique ou par l’analyse épistémologique des catégories scientifiques candidates au statut d’espèce naturelle (élément chimique, particules élémentaires, espèces biologiques, etc.). Au chapitre 4, je mets de l’avant un argument éliminativiste à l’égard de la notion d’espèce naturelle, c’est-à-dire que je démontre que le fait de centrer l’épistémologie de la classification sur un tel concept est plus nuisible qu’utile. Afin de le critiquer, il convient cependant de présenter le complexe discursif dont il est question, une tâche que j’entreprends dans ce chapitre-ci.

La tradition des espèces naturelles s’étant prouvée très prolifique au cours des dernières décennies, il m’est impossible, faute de temps et d’espace, de réaliser une prospection détaillée de chacune des approches qui composent le complexe discursif en question. La revue du corpus de textes que je suggère se veut donc plutôt un survol historique qui couvre une grande diversité d’approches des espèces naturelles, constitué autant de textes classiques que de travaux contemporains. Je profite de ce survol pour démontrer l’existence d’une structure épistémique inhérente à la tradition des espèces naturelles que j’appelle l’arche des espèces naturelles. En mettant à jour l’arche des espèces naturelles, ce chapitre-ci me permet de défendre, de concert avec Hacking (1991; 2007a) la thèse selon laquelle cette tradition est effectivement unitaire, du moins en philosophie des sciences. Mon argument du chapitre 4, qui mobilise les outils du

cohérentisme historique, consiste à démontrer que l'analyse épistémologique des pratiques classificatoires en sciences gagne à évacuer cette arche.

En bref, l'arche des espèces naturelles est composée des trois croyances suivantes, déclinées différemment par chaque théorie des espèces naturelles :

- 1- Les espèces naturelles sont des catégories métaphysiquement privilégiées;
- 2- Les espèces naturelles sont des catégories épistémiquement importantes;
- 3- Il faut se munir d'une théorie des espèces naturelles pour comprendre les pratiques classificatoires en sciences.

Ces croyances fondamentales pour la tradition des espèces naturelles sont ainsi liées : certaines catégories détiennent un statut métaphysique particulier (dont le détail varie d'une théorie à l'autre) qui explique (ou est dérivé de) leur spécificité épistémique, soit le fait qu'elles jouent un rôle primordial au sein des pratiques classificatoires; en conséquence, il faut se munir d'une ou de plusieurs théories des espèces naturelles afin de bien saisir la teneur des pratiques classificatoires en sciences. La troisième croyance s'appuie sur les deux premières. L'assemblage résultant forme ce que j'appelle l'arche des espèces naturelles (Figure 3.1).

Le chapitre est divisé comme suit. La première partie s'intéresse à trois importantes sources historiques de la tradition des espèces naturelles, soit l'Arbre de Porphyre, les travaux de Locke sur la notion d'essence et ceux des philosophes anglais du 19^e siècle (principalement Mill et Whewell) au travers desquels l'arche s'est cimentée. La deuxième partie s'intéresse aux approches s'inscrivant dans le corpus de textes étudié depuis la résurgence de la notion d'espèce naturelle dans la deuxième moitié du vingtième siècle. Dans le but de couvrir une partie significative du complexe discursif, je présente des approches essentialistes (Bird 2009, 2018; LaPorte 2004) et non essentialistes. Au nombre de ces dernières, j'appréhende les travaux de Dupré (1993) et de Boyd (1991, 1999a) avant d'illustrer la situation contemporaine par l'entremise de deux théories récentes et influentes, soit celles de Khalidi (2013) et de Slater (2015). Le tout me permet de défendre l'unité de la tradition des espèces naturelles et de mettre en lumière l'arche des espèces naturelles.

3.2 Les fondements historiques du complexe discursif

3.2.1 L'Arbre de Porphyre

La théorisation des espèces naturelles au sein de la tradition philosophique occidentale trouve sa source dans les écrits de Platon et d'Aristote. La notion d'espèce peut d'abord être perçue à la lumière du principe de division logique. Ce principe sous-tend le modèle arborescent qui devient, à la fin de l'Antiquité, l'Arbre de Porphyre. Au sein de ce modèle logique, trois notions s'inscrivent dans un rapport méréologique qui les définit : un *individu* fait partie d'une *espèce*, elle-même constitutive d'un *genre*. La relation espèce-genre, contrairement à celle individu-espèce, est transposable à différents niveaux d'organisation. Une classe d'objets peut être considérée comme une espèce par rapport à une classe plus inclusive, qui sera alors son genre d'appartenance. Elle peut aussi être considérée comme un genre par rapport à des classes moins inclusives, qui seront alors ses espèces constitutives. L'individu est le terme de l'analyse logique qui décompose un ensemble d'objets étudiés, alors que l'expression *summum genus* peut être utilisée pour référer à l'ensemble le plus inclusif. Il convient aussi de noter que les espèces se définissent par des caractéristiques qui leur sont spécifiques (d'où l'origine de l'adjectif), c'est-à-dire qui les distinguent des autres membres du genre. Ces caractéristiques, appelées *differentiae*, sont considérées comme des traits nécessaires et suffisants pour établir l'appartenance d'un individu à une espèce (Crombie 1994; Deleuze 1968; Panchen 1992; Zachos 2016).

Le principe de division logique est d'abord et avant tout un modèle permettant l'analyse d'un ensemble d'objets. À l'Antiquité du moins, sa portée ontologique est limitée. Chez Aristote, par exemple, le modèle qui rend compte des relations naturelles entre êtres biologiques est celui de l'échelle naturelle, de la progression linéaire, et non celui arborescent de l'embranchement et de la divergence qui correspond au principe de division logique (Lovejoy 1936; Rieppel 2010b; Winther 2012).

Le principe d'analyse mobilisant genre et espèce est donc un outil méthodologique d'analyse logique (Panchen 1992; Verboon 2014; Hacking 2007a), certes, mais il convient d'ajouter une précision. Les individus et les espèces les moins inclusives auxquelles ils appartiennent, c'est-à-dire les espèces terminales, sont bien considérés comme étant réels. Les

espèces terminales se distinguent des espèces plus inclusives par l'impossibilité de les décomposer en sous-groupes détenant chacun une essence propre, des *differentiae* qui seraient plus que de simples accidents (les accidents sont les particularités qui caractérisent les individus). Cette réification ontologique des espèces terminales prend tout son sens dans le cadre de la biologie des 18^e et 19^e siècles. En effet, alors que des débats portaient à l'époque sur la portée ontologique des schèmes classificatoires dans leur ensemble, la réalité métaphysique des espèces, considérées dès lors comme un échelon fixe de la hiérarchie, était au contraire largement acceptée (Hacking 2007a; Stevens 1997).

Cette brève présentation de l'Arbre de Porphyre permet de cerner trois objectifs que ce modèle vise à accomplir : décrire un individu sur la base des catégories successivement plus inclusives auxquelles il appartient (et non seulement sur la base d'une espèce d'appartenance), mettre en relation des catégories à différents niveaux d'organisation ainsi qu'identifier et qualifier des classes d'objets porteuses d'une valeur métaphysique notable, soit les espèces terminales. Le complexe discursif portant sur les espèces naturelles s'intéresse principalement à ce troisième axe et délaisse entièrement le second. L'objectif des théories des espèces naturelles, en philosophie des sciences, est en effet d'offrir une description ontologique ou épistémologique (et plus souvent les deux) de ces espèces qui permettent de rendre compte de l'ordre naturel, de ces espèces qui se distinguent d'autres catégories dites artificielles, non naturelles ou arbitraires.

3.2.2 Les essences lockéennes

L'un des développements théoriques importants de la tradition des espèces naturelles relève de la notion d'essence. Celle-ci, chez Aristote, réfère littéralement à ce qui fait d'une chose ce qu'elle est. Dans le cas des espèces, essence et *differentiae* vont de pair (en effet, les spécificités de l'espèce sont ces caractéristiques qui font d'elle un groupe distinct). Lorsqu'on parle de l'essence d'un individu plutôt que de celle d'une espèce, il s'agit de cet état (typique de son espèce) vers lequel il tend. Par exemple, bien qu'un nourrisson ne soit pas encore un être rationnel, toute son existence le fait tendre vers le développement de cette rationalité, considérée par Aristote (et bien d'autres) comme étant l'essence humaine.

L'une des importantes contributions de Locke (1632-1704), le premier auteur typiquement associé au complexe discursif portant sur les espèces naturelles, a été de sophistication la notion d'essence en théorisant la distinction entre essences réelles et essences nominales (Locke 1975 [1689]). L'essence réelle réfère à la teneur métaphysique d'une chose. Par exemple, l'essence réelle d'un objet matériel réfère aux causes physiques qui engendrent ces qualités observables. Cette essence réelle se rapproche du sens aristotélicien du terme, mais il faut souligner que ce sont exclusivement les individus qui détiennent une telle essence (Conn 2002; Crane 2003). Au contraire, l'essence nominale est l'idée qui réfère aux similarités que la cognition humaine peut identifier, par abstraction, dans un ensemble d'objets matériels (Conn 2002; Jones 2016).

Pour Locke, nous sommes incapables d'accéder aux essences réelles. Par contre, la connaissance peut être dite réelle si l'essence nominale des concepts tend à correspondre à l'essence réelle des membres types de la classe d'objets dont il est question (Phemister 1990; Stuart 1999). Cette lecture des travaux de Locke¹⁶ en fait l'un des pionniers du naturalisme, une posture qui défend notamment l'idée selon laquelle nos concepts scientifiques impliqués dans les pratiques classificatoires tendent (ou devraient tendre) à correspondre aux divisions qui existent dans la structure du monde naturel (Ereshefsky et Reydon 2015). Dans cette optique, une catégorie scientifique serait une espèce naturelle s'il s'agit d'une essence nominale qui renvoie à un ensemble d'objets détenant tous la même essence réelle et traduisant ainsi une division réelle dans la structure du monde.

Les théories contemporaines des espèces naturelles, qu'elles mobilisent la notion d'essence ou non, sont fermement ancrées dans cette logique lockéenne; le naturalisme domine

¹⁶ L'interprétation de Locke selon laquelle nos systèmes classificatoires sont composés d'essences nominales exclusivement, comme il le suggère au sujet des espèces biologiques et des systèmes taxonomiques, suggère un antiréalisme en opposition au naturalisme que je décris ici (Crane 2003; Jones 2016). Cette lecture a influencé des naturalistes des 18^e et 19^e siècles (Crombie 1994; Hoquet 2005; Tort 1989) et a été lourdement critiquée par les auteurs essentialistes de la deuxième moitié du 20^e siècle (Conn 2002; Crane 2003; Jones 2016). Les auteurs et autrices essentialistes ayant critiqué son antiréalisme (Kripke et Putnam, par exemple; Lowe 2011) ont tout de même contribué à développer des approches des espèces naturelles qui prônent l'arrimage des catégories scientifiques à l'essence des phénomènes naturels.

en effet ce paysage philosophique (Crane 2003). Bird, par exemple, rend explicite le lien entre naturalisme et réalisme à l'égard des espèces naturelles :

More commonly, an argument for weak realism is a corollary of an argument for naturalism. Thus one may argue for the existence of natural divisions by articulating an argument for scientific realism that focuses on the presence of classifications into kinds in science. Scientific realism tells us that modern science is our best guide to what things there are and what they are like. Furthermore, science is not merely our best guide, but also it is generally, when consensus has been reached, a good if not infallible guide to such things. Natural science does divide things into kinds—kinds of subatomic particle, kinds of chemical element or chemical compound, kinds of crystal structure, clades and possibly species, and other biological kinds, kinds of rock, kinds of star, and so forth. If such divisions are not natural then much science has got things massively wrong. Conversely, if science is by and large right, then we can have a high degree of confidence that the divisions science draws are genuinely natural. (Bird 2018, 1399)

Comme je le démontre plus loin, même les approches qu'on pourrait associer à un très faible réalisme par rapport aux espèces naturelles (Boyd 1999b; Dupré 1993; Ereshefsky et Reydon 2015; Slater 2015) conservent l'idée selon laquelle il y a un lien intime entre espèces naturelles, succès scientifique et la structure du monde.

3.2.3 La cristallisation du schème de pensée relié aux espèces naturelles

Le succès scientifique auquel les approches contemporaines réfèrent est souvent formulé en termes inductifs, ce qui nous amène à la troisième source historique qu'il faut considérer afin de comprendre le complexe discursif à l'étude. L'association de cette qualité inférentielle avec les espèces naturelles renvoie aux écrits des auteurs anglais William Whewell (1794-1866) et John Stuart Mill (1806-1873). Ces auteurs sont, selon Hacking (1991; 2007a), les fondateurs de la tradition des espèces naturelles.

Whewell (1840) entendait contribuer à la discussion pérenne portant sur le statut métaphysique des différents rangs classificatoires de la systématique biologique (débat mentionnés plus tôt), ce qui l'a amené à théoriser la notion de *classe naturelle*. Les classes naturelles représentent les catégories scientifiques qui peuvent faire l'objet d'inductions précises (Hacking 2007a). Or, ultimement, cette qualité des classes naturelles découle de leur teneur ontologique : « what makes kinds natural for Whewell is ultimately the 'constitution' and 'construction' of things. » (Magnus 2015b, 272).

Mill, quant à lui, introduit la notion d'espèce réelle (*real Kind*), qui réfère à des groupes d'entités détenant un statut métaphysique particulier, afin de les distinguer des catégories épistémiquement utiles, mais qui sont privées d'un tel statut (Mill 1862). Contrairement à Whewell et aux approches d'inspiration lockéenne, Mill évite cependant d'ancrer le statut métaphysique particulier des espèces naturelles dans une essence réelle ou dans la « constitution » ontologique des classes d'objets (ou de leurs occurrences). Selon lui, les espèces réelles sont des catégories au sujet desquelles un nombre illimité d'inférences inductives peuvent être faites. Cette caractéristique épistémique les distingue d'autres catégories qui sont tout à fait utiles aux pratiques scientifiques en raison des liens de similarité qui unissent leurs membres, mais pour lesquels le nombre de liens de similarité est limité plutôt qu'infini (Hacking 2007a; Magnus 2014c; 2015b).

Mill explique ce trait spécifique des espèces réelles par un postulat métaphysique : elles engendrent la régularité de sorte à soutenir un nombre illimité d'inférences inductives au même titre qu'une même cause engendre toujours les mêmes effets, toutes choses étant égales par ailleurs; cependant, la régularité associée aux espèces réelles est synchronique alors que celle associée aux causes est diachronique. Dans cette optique, le fait même d'appartenir à une espèce réelle explique pourquoi des individus partagent une infinité de traits caractéristiques qui rendent possibles une infinité d'inférences inductives. Cette situation est synchronique en ce sens que l'appartenance à l'espèce réelle et les traits caractéristiques permettant l'induction existent simultanément (contrairement à la cause et à l'effet qui se suivent dans le temps; Magnus 2014b, 2015b; Mill 1862).

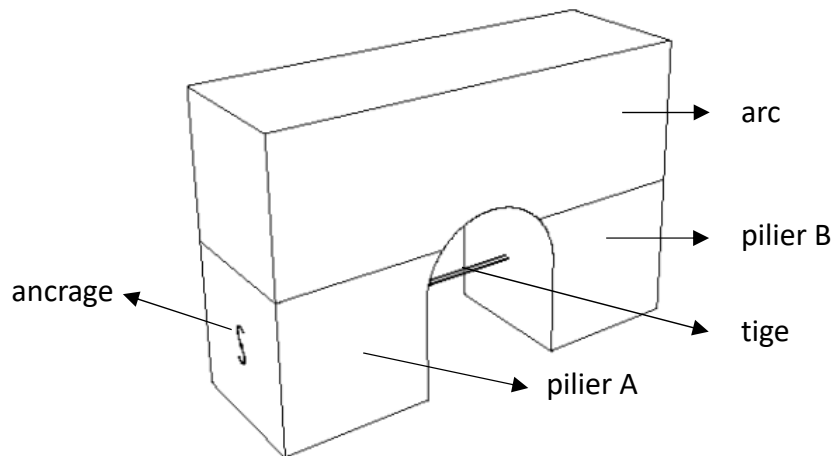
Ainsi, des catégories non réelles peuvent certainement, selon Mill, être utiles à des fins inférentielles ou taxonomiques. Cependant, les projets de classification doivent chercher à se conformer aux espèces réelles et ainsi à découper le monde en fonction de ses articulations qui existent indépendamment de la cognition humaine (*to carve the world at its joints*) :

Since the common properties of a true Kind, and consequently the general assertions which can be made respecting it, or which are certain to be made hereafter as our knowledge extends, are indefinite and inexhaustible; and since the very first principle of natural classification is that of forming the classes so that the objects composing each may have the greatest number of properties in common; this principle prescribes

that every such classification shall recognize and adopt into itself all distinctions of Kind, which exists among the objects it professes to classify. Accordingly, all natural arrangements, whether the reality of the distinction of Kinds was felt or not by their framers, have been led, by the mere pursuit of their own proper end, to conform themselves to the distinctions of Kind. (Mill 1862, 438)

Un ensemble de croyances s'est donc bel et bien cristallisé chez les auteurs anglais du 19^e siècle. Mill, Whewell ainsi que leurs successeurs et successeuses (notamment John Venn [1834-1923], à qui on doit l'expression « *natural kind* ») accordent une importance marquée au rapport entre induction, classification et métaphysique. Le fait que les espèces naturelles soient des choses ontologiques réelles, des éléments de la structure du monde, des entités métaphysiques et non de simples outils épistémiques (des outils mnémotechniques, par exemple), est inscrit au cœur de leurs philosophies. Dans les travaux de Mill et de Whewell, il est clair que ce constat métaphysique explique au moins en partie les succès des pratiques scientifiques. Cela mène à une réitération à peine modifiée de la convergence des essences nominales sur les essences réelles : les pratiques scientifiques sont efficaces parce qu'elles savent s'arrimer à la structure du monde et rendre compte des divisions naturelles qui y figurent, que ces auteurs considéraient comme résultant de l'œuvre du Créateur. Il est donc tout naturel de vouloir théoriser épistémologiquement ces catégories scientifiques toutes particulières, ce qui devrait illuminer les spécificités des pratiques classificatoires. Je propose finalement de représenter les croyances extraites par abstraction de ces travaux fondateurs de la tradition des espèces naturelles sous la forme d'une arche (Figure 3.1).

Figure 3.1 – L'arche des espèces naturelles



Cette figure représente un ensemble de trois croyances inhérentes à la tradition des espèces naturelles et les relations logiques liant ces trois croyances. L'arc est la croyance selon laquelle il est nécessaire de se doter d'une ou de plusieurs théories des espèces naturelles afin d'appréhender les pratiques classificatoires d'un point de vue épistémologique. Cette croyance s'appuie sur deux piliers. Le pilier A représente la croyance selon laquelle les espèces naturelles sont des catégories scientifiques détenant un statut métaphysique particulier. Le pilier B représente la croyance selon laquelle elles détiennent aussi un statut épistémologique particulier. La stabilité de la structure est assurée par un ancrage en esse (deux ancrages et une tige les reliant) qui représente la croyance selon laquelle le statut métaphysique des espèces naturelles peut être déduit de leur statut épistémologique ou que leur statut métaphysique (alors traduit par des constats ontologiques à leur sujet) explique leur statut épistémologique.

L'arc de cette structure est formé par la croyance selon laquelle il est important pour la philosophie des sciences de se doter d'une ou de plusieurs théories des espèces naturelles pour comprendre les pratiques classificatoires. Cet arc repose sur deux piliers. Le premier consiste en la croyance selon laquelle *certaines* catégories scientifiques détiennent un statut métaphysique privilégié, soit celui d'espèce naturelle qui les démarque du reste, décrite ontologiquement par des concepts comme ceux d'essence, de constitution, de régularité synchronique, etc. Le second pilier consiste en la croyance selon laquelle les espèces naturelles jouent un rôle crucial au sein des inférences inductives ou des explications scientifiques. La stabilité de ces deux piliers et de la structure dans son ensemble est garantie par un ancrage en esse qui lie les deux piliers de sorte à empêcher leur éloignement. Cette esse est la croyance selon laquelle le fait métaphysique (les caractéristiques ontologiques des catégories dites espèces naturelles) vient expliquer le fait épistémologique; elle prend parfois une forme inverse, soit la croyance selon laquelle le fait épistémologique permet de déduire un fait métaphysique au sujet de catégories scientifiques. L'omniprésence de ces piliers et de cette tige les solidifiant au sein des textes contemporains est

parfaitement traduite par la citation suivante : « This much is common ground among essentialists and nonessentialists. Indeed, nonessentialists philosophers also maintain that natural kinds are discoverable by science and constitute the basis for projection and inductive inference. » (Khalidi 2013, 14).

En bref, je suggère que toute théorie des espèces naturelles s'appuie sur ces deux piliers et le lien qui les unit. C'est en effet ce qui justifie le besoin de développer un propos philosophique au sujet des espèces naturelles : il faut théoriser les espèces naturelles *parce que* celles-ci détiennent un statut métaphysique et épistémique particulier. Dans cette optique, comprendre les pratiques scientifiques liées à la classification, c'est bien comprendre ces catégories toutes particulières. Dans les pages qui suivent, j'illustre comment les approches plus récentes déclinent l'arche des espèces naturelles. L'omniprésence de l'arche soutient ainsi la thèse selon laquelle la tradition des espèces naturelle forme un complexe discursif unifié par cet ensemble de croyances fondatrices¹⁷. De surcroît, cette omniprésence fait de l'arche un tertre de départ adéquat pour propulser mon argument éliminativiste détaillé au chapitre 4.

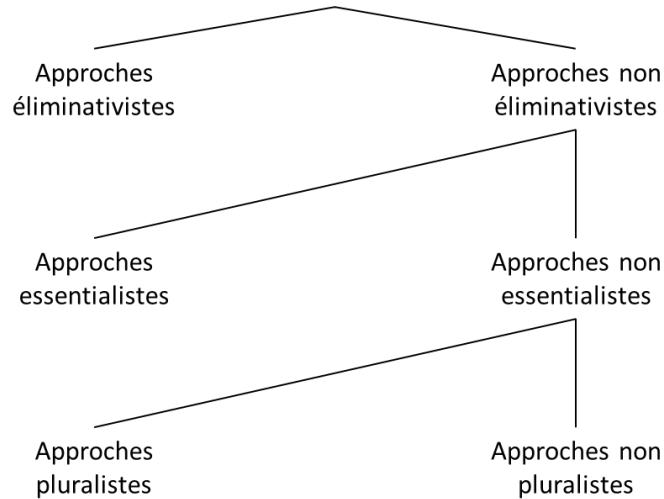
3.3 Les théories contemporaines des espèces naturelles

Je propose d'utiliser l'Arbre de Porphyre comme modèle pour broser le paysage contemporain relatif aux espèces naturelles en philosophie des sciences (Figure 3.2). De cette façon, on peut distinguer les approches éliminativistes et les approches non éliminativistes. Au sein des approches non éliminativistes, on peut distinguer les approches essentialistes et non essentialistes. Au sein de ces dernières, on peut finalement distinguer les approches pluralistes des non pluralistes. Cette description est approximative et néglige, comme tout schème classificatoire, l'hétérogénéité qui habite inévitablement chacune des catégories mobilisées. De surcroît, il faut préciser que d'autres façons de découper le complexe discursif relatif aux espèces

¹⁷ La posture de Hacking selon laquelle il y aurait *une* tradition des espèces naturelles a été critiquée. Cependant, je me permets de suivre Hacking sur ce point pour deux raisons. Premièrement, je m'intéresse seulement aux théories des espèces naturelles en philosophie des sciences. J'exclus donc de fait une partie de l'hétérogénéité irréductible (issue de la philosophie de l'esprit et du langage) du corpus de textes qui a servi à remettre en question l'unicité de la tradition. Deuxièmement, je démontre dans ce chapitre que les propos au sujet des espèces naturelles en philosophie des sciences témoignent effectivement d'une unicité. Une fois les spécificités abstraites, toutes ces approches présentent des éléments constitutifs partagés et cruciaux, soit ceux représentés par l'arche.

naturelles sont valables. Nonobstant ces limitations, ce portrait permet tout de même d'appréhender le corpus de textes de façon à en identifier les principales tendances.

Figure 3.2 – Les principales approches contemporaines constitutives de la tradition des espèces naturelles



Dans ce qui suit, j'explore chacune des approches non éliminativistes en me concentrant sur certains textes influents (les approches éliminativistes sont présentées au prochain chapitre). Considérant l'étendue du complexe discursif à l'étude, il est impossible de donner le détail de chacune des théories des espèces naturelles qui ont été développées dans les dernières décennies. Le survol ici effectué permet tout de même d'identifier la présence marquée, au sein des théories des espèces naturelles contemporaines, des croyances identifiées dans la section précédente, soit celles qui forment l'arche des espèces naturelles. En 3.3.1, j'illustre comment les approches essentialistes déclinent ces croyances. En 3.3.2, je fais le même travail pour les approches pluralistes, avant de réitérer la manœuvre, en 3.3.3, pour les approches monistes, mais non essentialistes.

3.3.1 Les approches essentialistes

Après que la notion d'espèce naturelle a été partiellement ressuscitée par Russell (1872-1970) au milieu du 20^e siècle (Russell 1948; Hacking 1991), ce sont finalement les travaux de Kripke (1972) et de Putnam (1973, 1990) qui l'ancrent solidement au sein de la philosophie des sciences occidentales et qui donnent ainsi le coup d'envoi aux réflexions contemporaines sur le

sujet. Le haut fait d'armes de ces auteurs (voir Hacking 2007b pour les distinctions entre les deux approches) a été de réhabiliter une conception essentialiste dans l'analyse de concepts théoriques comme ceux typiquement associés à la notion d'espèce naturelle.

L'essentialisme, chez Aristote, stipule que les choses tendent vers un but prédéterminé, leur l'essence (par exemple, les êtres humains tendent à développer leurs capacités rationnelles). Cette version de l'essentialisme est dite téléologique (le terme *telos* signifiant « but » ou « fin » en grec ancien) puisqu'elle insiste sur la propension des choses à réaliser une essence qui les précéderait et qui déterminerait leur développement (le cas échéant). Cette approche a été fortement critiquée au cours du 20^e siècle, notamment par les empiristes logiques, Quine et Popper (Devitt 2008; Wilkerson 1988). Or, ces critiques reposaient plutôt sur l'aspect téléologique de cette doctrine que sur l'idée que les choses détiennent une essence propre. Ainsi, lorsque Putnam et Kripke articulent des conceptions essentialistes des espèces naturelles, ils le font en évitant toute référence à la nature téléologique des choses. Je décris sous peu le détail des postures essentialistes de la fin du 20^e siècle.

Les approches essentialistes plus récentes conservent l'esprit instauré par les travaux de ces deux auteurs tout en répondant à certaines difficultés que leur conception essentialiste des espèces naturelles a introduites en philosophie des sciences. Il faut tout de même noter que les travaux de Kripke et Putnam marquent une rupture au cœur de « la » tradition des espèces naturelles. En effet, à partir de ce moment, les espèces naturelles ont été considérées, d'une part, comme de potentiels universaux dans une perspective de philosophie du langage (Devitt 2007, 2011; Fernández Moreno 2016; Wikforss 2005; Häggqvist et Wikforss 2017) et d'autre part, dans la perspective de philosophie des sciences qui m'intéresse.

Ces considérations historiques étant établies, il convient de se pencher sur l'essence des théories essentialistes. L'idée centrale est celle d'un fait ontologique inhérent à chacune des occurrences d'une espèce naturelle, soit son essence, qui serait apte à expliquer une grande partie des caractéristiques associées à la classe d'objets en question (Bird 2009, 2018; Devitt 2008; Ellis 2001; Havstad 2017; Soames 2002; Wilkerson 1988, 1993). Sous l'influence de Kripke et de Putnam, ce fait ontologique est typiquement considéré, dans une perspective

réductionniste, comme relevant de la microstructure physico-chimique intrinsèque aux entités concernées. Le nombre de protons présents dans le noyau d'un atome est, dans cette perspective, un exemple paradigmatique d'essence et un élément chimique est alors un exemple paradigmatique d'espèce naturelle (Havstad 2017). Le bagage génétique d'individus biologiques a aussi été considéré comme potentielle essence pour des populations (ou des espèces biologiques) qui seraient ainsi considérées comme des espèces naturelles (Devitt 2008).

Afin de circonscrire rigoureusement la teneur de la notion d'essence dans le cadre de la tradition des espèces naturelles, trois critères définitionnels ont été suggérés. Les travaux de Kripke et de Putnam ont érigé l'essentialisme autour d'un premier critère, soit celui de la nécessité modale. Dans cette perspective, l'essence est une caractéristique nécessaire et suffisante pour l'appartenance d'un individu à une espèce naturelle. Cette prémisse est à interpréter d'une manière bien spécifique (Khalidi 2013) : si un individu appartient à une espèce naturelle, il y appartient dans tous les mondes possibles. Autrement dit, le fait d'avoir certaines caractéristiques essentielles détermine absolument cet individu et son appartenance à une espèce naturelle. En ce sens, les termes associés à une espèce naturelle, comme « eau » ou « or » sont considérés comme étant des référents rigides (*rigid designator*).

Depuis les travaux de Putnam et de Kripke, l'importance de la nécessité modale a été remise en question au sein même de la mouvance essentialiste. En effet, si on doit comprendre le fait que les termes associés aux espèces naturelles « réfèrent aux mêmes entités » de manière stricte, alors cette contrainte est tout simplement fautive. Les tigres qui existent dans notre monde sont certainement des individus différents de ceux qui existeraient dans un autre monde (hypothétique). Si, de manière plus charitable, on entend par « mêmes entités » les entités qui partagent les mêmes caractéristiques, alors le critère de la nécessité modale rend l'étiquette « espèce naturelle » triviale. En effet, tout nom commun avec une définition plus ou moins précise pourrait, sur cette base, aspirer au statut d'espèce naturelle. Or, l'essentialisme tend plutôt à maintenir l'injonction à distinguer les espèces naturelles des catégories non naturelles, injonction inhérente à l'arche des espèces naturelles. Bird insiste sur ce point : « Kripke treated natural kind terms along with singular proper names as regards rigid designation, but supplied no

account of what precisely this could mean for a natural kind term. The worry is that if any sense can be made of this at all, it leads to triviality » (Bird 2018, 1414).

En raison de cette difficulté relative à la nécessité modale, les approches essentialistes plus récentes se sont recentrées sur deux autres critères (déjà présents dans les travaux de Kripke et de Putnam) pour l'identification et la description des espèces naturelles : elles doivent détenir une essence qui 1) est *intrinsèque* aux représentants de la classe et 2) relève de caractéristiques *microstructurales* (Bird 2010; Devitt 2008; Ellis 2001; Havstad 2017; Kripke 1972; LaPorte 2004; Putnam 1973). Dans ce qui suit, je me concentre dans un premier temps sur le critère selon lequel l'essence doit être intrinsèque aux individus, critère que j'explore par l'entremise de l'approche essentialiste de Devitt à l'égard des espèces biologiques (Devitt 2008). Plus loin, j'appréhende le critère microstructural par l'analyse d'un débat entre LaPorte (2004) et Bird (2010), deux autres auteurs associés à l'essentialisme. Je profite de l'occasion pour mentionner des points faibles de l'essentialisme. Ces légers détours permettent de mieux saisir les écueils qui ont motivé des philosophes à formuler d'autres approches des espèces naturelles.

3.3.1.1 L'essence intrinsèque

L'argument de Devitt vise à présenter les espèces biologiques comme des espèces naturelles identifiées sur la base d'une essence au moins partiellement intrinsèque. Son argument repose sur une distinction importante que le regard traditionnel sur les catégories biologiques néglige, selon lui, depuis Darwin. Cette distinction concerne deux niveaux d'explication. Dans la perspective traditionnelle, les espèces biologiques sont définies par les traits relationnels des individus qui les composent, soit l'ascendance commune des représentants de l'espèce. Cette idée renvoie à l'identification du statut d'un groupe d'individus puisqu'elle permet de dire la chose suivante : « ces individus forment une espèce biologique puisqu'ils forment une communauté d'ascendance commune ». Au contraire, ce critère ne permet aucunement, toujours selon Devitt, de savoir si un organisme donné fait bien partie d'une espèce biologique donnée. Les traits intrinsèques des organismes, en contraste, permettent de déterminer une telle appartenance. En ce sens, nous dit Devitt, l'essence ou la nature d'une espèce réfère aux traits intrinsèques et principalement génétiques qui permettent de caractériser les liens de similarité entre organismes. Devitt insiste sur le fait que sa posture est complémentaire avec celle

considérant les espèces comme des entités historiques, ce qui explique qu'il considère l'essence comme étant « au moins partiellement » (*at least partially*) intrinsèque.

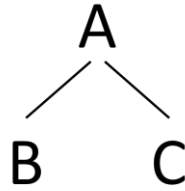
L'argument de Devitt a le mérite d'être très détaillé et d'appréhender la question de l'ontologie des espèces biologiques sous divers angles et selon diverses définitions de la notion. Je focalise cependant mon attention sur la partie de son argument qui permet d'expliquer pourquoi les caractéristiques *intrinsèques* des espèces devraient être incluses dans leur essence, un argument qui s'oppose à d'autres perspectives essentialistes au sujet des espèces biologiques (Griffiths 1999; Okasha 2002) ainsi qu'à la perspective traditionnelle sur la question (Ghiselin 1974; Hull 1976; Zachos 2016). Cet argument repose sur la capacité des essences intrinsèques à expliquer le potentiel de généralisation et d'explication associé aux espèces biologiques. Devitt avance la chose suivante :

Generalizations of this kind demand an explanation. *Why are they so?* [...] Explanations will make some appeal to the environment, but they cannot appeal only to that. There has to be something about the very nature of the group — a group that appears to be a species or taxon of some other sort — that, given its environment, determines the truth of the generalization. That something is an intrinsic underlying, probably largely genetic, property that is part of the essence of the group. Indeed, what else could it be? (Devitt 2008, 352)

Devitt étoffe notamment cet argument en démontrant que l'approche phylogénétique-cladiste de la notion d'espèce est incapable de déterminer avec précision l'extension d'une espèce biologique. Le concept phylogénétique-cladiste de l'espèce, selon Devitt, suggère qu'une espèce est un segment de l'Arbre du vivant contenu entre deux embranchements. Chaque fois qu'il y a événement de spéciation, une espèce mère laisse donc place à deux espèces filles.

Selon l'approche phylogénétique-cladiste, la seule façon de déterminer si un individu appartient à une espèce implique de savoir s'il fait partie de la lignée de descendance à laquelle une espèce comme « tigre » réfère. Devitt nous invite à considérer un arbre représentant les relations entre sept espèces afin de démontrer que cette détermination phylogénétique de l'appartenance d'un individu à une lignée est faible d'un point de vue explicatif. Je simplifie ici en prenant simplement les trois premières espèces de l'arbre, soit A, B et C; A est l'ancêtre commun de B et C.

Figure 3.3 – Un arbre phylogénétique formé de trois espèces



Selon Devitt, la seule information fournie par un tel arbre au sujet de l'espèce B est que celle-ci descend de A et se distingue de C. Aucune information ne serait ainsi fournie concernant la nature de chacun des taxons : on sait que B et C sont distincts, mais il demeure impossible de déterminer si un individu donné appartient à B ou à C (dans la citation qui suit, l'acronyme « P-CC » réfère au concept phylogénétique-cladiste d'espèce) : « Relational facts of the sort [...], which are the only facts that P-CC allows to constitute the natures of species, would do nothing to distinguish lions from tigers, hence nothing to explain the morphological, physiological, and behavioral differences between them. » (Devitt 2008, 368).

Le problème avec la logique de Devitt est qu'il s'intéresse *au produit fini* des pratiques phylogénétiques sans considérer le reste des pratiques menant à la reconstruction d'arbres. Celles-ci mobilisent un ensemble de données au sujet d'unités taxonomiques opérationnelles, dans certains cas des espèces biologiques, afin de caractériser les liens de similarité entre elles. Cette caractérisation des liens de similarité est ensuite mobilisée pour inférer des relations phylogénétiques (Felsenstein 2004; Wiley et Lieberman 2011). C'est donc dire que lorsqu'un arbre est reconstruit, celui-ci ne nous informe aucunement sur l'appartenance des individus (gènes, organismes, etc.) aux unités taxonomiques (familles de gènes, espèces, genres, etc.) parce que ce travail est effectué en amont, habituellement sur la base de critères de similarité variés (similarité morphologique, génétique, physiologique et comportementale). La phylogénétique nous informe plutôt sur les relations entre les unités taxonomiques étudiées. Ce fait est occulté par une analyse philosophique se concentrant exclusivement sur le produit fini des pratiques classificatoires et tout particulièrement sur les catégories constitutives de ce produit fini.

Devitt pourrait répondre à cela qu'il faut conséquemment redéfinir la notion d'espèce phylogénétique-cladiste afin qu'elle honore la similarité en question. Or, ce geste rhétorique viendrait nuire au projet de la phylogénétique traditionnelle qui cherche notamment à exclure de ses descriptions des groupes d'organismes constitués par une ressemblance qui serait due à des facteurs autres que la descendance commune. La similarité est pertinente pour la phylogénétique si et seulement si elle est causée (expliquée) par l'ascendance commune. En cohérence avec cette croyance, les praticiens et praticiennes qui adhèrent au patron arborescent construisent des modèles afin d'exclure les similarités qui seraient dues à des événements réticulaires, comme la convergence de deux lignées. En ce sens, l'argument de Devitt selon lequel les traits intrinsèques des organismes permettent d'expliquer l'appartenance d'un individu à une espèce (ou la nature de l'espèce) alors que les traits relationnels ou extrinsèques en sont incapables est inexact.

La thèse de Devitt fait face à un problème additionnel qui relève du fait même de considérer les traits génétiques comme intrinsèques¹⁸ aux organismes (il accepte qu'aucun gène ne soit nécessaire et suffisant à l'appartenance à une espèce, mais il considère que des groupes de gènes [*clusters*] ou des programmes génétiques peuvent servir d'essence pour une espèce). Il a pourtant été démontré à moultes reprises que les gènes, modules développementaux et génomes réalisent leurs fonctions biologiques au sein d'un organisme si et seulement si des conditions environnementales, à concevoir à différentes échelles, sont réalisées (Dupré 2012; Halfon 2017; Keller 2014; McClintock 1984; Niewöhner et Lock 2018). Ce point est d'autant plus important dans le contexte actuel qui est celui de la reconnaissance grandissante de l'importance explicative, notamment d'un point de vue évolutif, des interactions interorganismes et interespèces. Par exemple, dans le monde microbien, il est commun de voir la fonction d'un gène déterminé par les interactions d'un microbe avec sa communauté (Baptiste et Papale 2021; Bernard et al. 2020; Doolittle et Inkpen 2018; Papale, Saget et Baptiste 2020; Whiteley, Diggle et Greenberg 2017). De plus, beaucoup de gènes sont mobiles, ayant la capacité de se déplacer d'un génome à l'autre (ou au sein d'un génome), de sorte qu'il soit problématique de les considérer comme des traits

¹⁸ Par ailleurs, l'argument de Devitt est vulnérable aux différentes critiques portant sur la signification même du terme « intrinsèque » (voir Khalidi 2013, 31-33). Je me concentre cependant sur les aspects de son approche qui relèvent plus directement de la philosophie de la biologie.

intrinsèques des organismes qui les portent (Doolittle 1999, 2000; Ghaly et Gillings 2018; Koonin, Makarova et Wolf 2021; McDaniel et al. 2010; Bapteste et Boucher 2008). Finalement, même les gènes ou ensembles de gènes qui demeurent sagement au sein d'un génome tendent à avoir des origines phylogénétiques plurielles (Kawashima et al. 2009; Méheust, Lopez et Bapteste 2015; Méheust et al. 2018; Papale, Saget et Bapteste 2020; Pasek, Risler et Brezellec 2006; de Souza 2012). L'entrelacement fonctionnel irréductible entre microbes et macrobe implique que le tout soit valide pour l'un autant que pour l'autre (Campbell et al. 2020; Chiu et Gilbert 2015; Hu et al. 2016; Hu, Gao et Zhu 2017; Rosenberg et Zilber-Rosenberg 2018).

Face à ces constats biologiques, il convient d'admettre que la focalisation sur des gènes « intrinsèques » pour définir la nature d'une espèce relève d'une pétition de principe (focaliser la recherche sur des gènes détenant des lignées stables dont les histoires évolutives concordent [les marqueurs génétiques traditionnels] est utile parce que nous cherchons des lignées stables et concordantes). L'utilisation de codes-barres (petites séquences d'ADN peu sujettes aux mutations) pour identifier des espèces témoigne du fait que cette pétition de principe, loin d'être le fait des théories essentialistes, est bien à l'œuvre au sein des pratiques taxonomiques contemporaines (Hobern 2021). Or, la génétique perd une partie significative de son pouvoir explicatif relatif aux traits comportementaux, morphologiques, physiologiques et même phylogénétiques si elle fait abstraction des interactions interindividus qui déterminent ces phénomènes. En ce sens, l'argument de Devitt est ancré dans une conception weismanienne du matériel génétique et de son rôle au sein du monde biologique. Cela nuit à la bonne compréhension des phénomènes génétiques en imposant des limites aux individus biologiques, des barrières dont la validité empirique est douteuse.

Les critiques précédentes ont été formulées à l'égard des travaux de Devitt, mais s'appliquent, *mutatis mutandis*, aux catégories physiques et chimiques. Par exemple, Khalidi (2013) démontre que des particules fondamentales, comme les quarks, sont caractérisées par des traits relationnels, comme le fait d'être lié à au moins un autre quark. Quant à elle, Havstadt (2017) souligne le besoin pressant de raffiner l'essentialisme pour rendre compte des catégories chimiques dont le comportement est largement déterminé par le contexte. L'accent essentialiste parfois mis sur les caractéristiques *intrinsèques* prête donc le flanc à diverses critiques qui

dépassent le contexte de la classification biologique et qui ont contribué à faire émerger des approches non essentialistes aux espèces naturelles.

3.3.1.2 L'essence microstructurale

Considérant ces problèmes associés à l'essence intrinsèque, il pourrait être plus judicieux de définir la notion d'essence en se fiant au réductionnisme qui guide notre attention vers des caractéristiques microstructurales. Un échange entre LaPorte et Bird permet d'illustrer la chose et ses problèmes. Dans sa réponse à LaPorte (2004), qui suggère que la microstructure d'une substance comme l'eau est insuffisante pour déterminer la teneur de l'espèce naturelle concernée, Bird (2010) insiste au contraire sur l'arrimage de l'essentialisme microstructural avec les pratiques scientifiques.

LaPorte argumente que la catégorie « eau », utilisée comme exemple et caractérisée exclusivement sur la base de sa composition chimique, pourrait être fractionnée. Concrètement, les molécules d' H_2O peuvent prendre différentes formes isotopiques. L'une d'entre elles contient du deutérium, un isotope d'hydrogène présentant un neutron (un des rares isotopes ayant un nom qui lui est propre), alors que la forme la plus commune d' H_2O est formée d'atomes d'hydrogène sans neutron, un isotope appelé protium.

Dans une expérience de pensée (comparable à celle des terres jumelles), LaPorte entrevoit la possibilité de considérer ces deux types d'eau (l'eau lourde [D_2O], formée par du deutérium, et l'eau formée de protium [P_2O], présente en beaucoup plus grande quantité dans l'atmosphère terrestre) comme relevant de catégories chimiques distinctes et donc d'espèces naturelles différentes. En effet, les deux molécules ont des comportements différents à plusieurs égards. Par exemple, la consommation d'eau lourde est néfaste pour les organismes (elle a un impact sur la mitose et endommage ainsi différents tissus au fil du temps). Or, ces effets de l'eau lourde sont dus à la structure des atomes concernés et plus spécifiquement à la présence de neutrons dans les atomes d'hydrogène d'eau lourde. Simplement en gardant l'accent sur la microstructure, on devrait donc diviser la catégorie « H_2O » en deux, mais les praticiens et praticiennes, nous dit LaPorte, prennent la *décision* de conserver la catégorie « H_2O » intacte. Peu importe ce qui influence cette décision, il semble qu'elle soit sous-déterminée par la microstructure.

La réponse de Bird à ce sujet revient à faire l'autruche. Selon lui, les différentes molécules d' H_2O doivent indéniablement être classées ensemble, de sorte que les praticiens et praticiennes seraient privé-e-s du pouvoir décisionnel que LaPorte leur reconnaît. Cette contrainte, Bird l'ancre dans la structure *moléculaire* d' H_2O qui demeure la même, peu importe que la masse atomique de ses atomes constitutifs change ou non (Bird s'appuie sur l'autorité de Linus Pauling pour prioriser ce critère). Il y a très certainement des différences entre les isotopes, mais ces différences, nous dit Bird, sont quantitatives et peuvent donc être négligées. Dans les faits, Bird priorise certaines caractéristiques microstructurales au détriment d'autres caractéristiques :

D_2O differs chemically from the other isotopic variants as regards certain quantitative features, such as rates of reaction. Strictly, all isotopic variants of all chemical substances differ from one another in such quantitative ways, but the difference is much more marked in the case of D_2O . As a result of this marked difference in reaction rate, pure D_2O in place of water can be poisonous for many organisms. LaPorte regards this as one of the reasons why scientists in his twin-earth story would be willing to exclude D_2O from the extension of *water_N*. Another reason is that D_2O can be used in the manufacture of fusion bombs in a way that isotopic variants cannot. Note that both these reasons come from outside chemistry. They are reasons, therefore, that are not pertinent to the science whose job is to investigate the nature of and to classify water. (Bird 2010, 127-128)

L'argument de Bird repose sur le fait que certaines considérations microstructurales peuvent être exclues sur la base des frontières disciplinaires qui structurent le milieu académique occidental. En effet, il considère *de facto* que la chimie est la science qui doit guider l'identification des espèces naturelles moléculaires. Puisque les différences entre isotopes n'ont aucune influence sur la charge électromagnétique des molécules, il est tout à fait adéquat de considérer que ces différences sont à l'arrière-plan des caractéristiques moléculaires typiquement étudiées par la chimie; la différence principale (et non la seule) entre les isotopes de la plupart des atomes étant leur taux de désintégration, leur étude relève principalement de la physique. Il n'en demeure pas moins qu'il faut ici faire un choix entre différents cadres théoriques pour savoir quelle est l'extension « naturelle » du terme « eau. » Bird ne fournit aucun argument pour justifier sa croyance selon laquelle la microstructure dictée par la chimie serait plus « naturelle » que celle dictée par l'étude des isotopes; il tient la chose pour acquise, probablement parce qu'il considère de manière apriorique, sous l'influence de Putnam et de Kripke, que le terme « eau » renvoie à une structure moléculaire spécifique. En ce sens, il s'agit d'une pétition de principe qui

est soulignée efficacement par l'argument de LaPorte (qui rappelle le problème de la coordination) : une décision, souvent prise implicitement plutôt qu'explicitement, vient déterminer quelles caractéristiques microstructurales, parmi des candidates plausibles et utiles, définissent une espèce naturelle.

En résumé, l'essentialisme, tel qu'il est illustré par le parti de Bird dans l'argument qui précède, mobilise l'idée d'une essence microstructurale dans l'espoir que celle-ci puisse déterminer l'extension d'une espèce naturelle de manière univoque, à la manière d'un référent rigide, sans pour autant trivialisier la notion d'espèce naturelle (rares sont les noms communs pouvant être associés à des microstructures de manière univoque). LaPorte argumente cependant que les catégories chimiques sont sous-déterminées par leur microstructure, un argument qui est accepté de façon consensuelle relativement aux catégories biologiques (Boyd 1999a; Brigandt 2009; Ereshefsky et Reydon 2015; Zachos 2016). Khalidi (2013, 37-41) démontre quant à lui que le microstructuralisme est généralement problématique puisqu'il implique l'abandon de toute causalité suggérée à des niveaux d'organisation supérieurs à celui des particules fondamentales de la physique (incluant le niveau des atomes et de leurs constituants, les protons, neutrons et électrons).

3.3.1.3 L'essentialisme, en bref

L'essentialisme au sujet des espèces naturelles consiste à associer des catégories scientifiques avec des essences typiquement considérées comme étant intrinsèques aux individus et relevant de faits microstructuraux. Les auteurs auxquels j'ai porté plus d'attention (Devitt et Bird) considèrent que la notion d'essence explique l'appartenance d'un individu à une catégorie (et la délimitation des frontières d'une espèce naturelle) ainsi que le pouvoir explicatif des espèces naturelles, typiquement considéré comme relevant des généralisations rendues possibles par l'existence de ces espèces naturelles. Les trois croyances s'étant cristallisées dans les textes classiques de la tradition des espèces naturelles sont déclinées avec une saveur propre à chacun des auteurs, mais elles sont tout de même bien présentes : 1) certaines catégories détiennent un statut métaphysique particulier, décrit en termes d'essences; 2) ce trait métaphysique, typiquement microstructural et intrinsèque aux représentants d'une espèce naturelle, est la source du pouvoir explicatif ou inductif des espèces naturelles et des succès

épistémiques des pratiques les mobilisant; 3) une théorie essentialiste des espèces naturelles permet de rendre compte du tout.

3.3.2 Les approches pluralistes

Les approches non essentialistes des espèces naturelles sont multiples et forment un ensemble hétérogène. Dès que les travaux de Kripke et de Putnam ont attiré le regard des philosophes des sciences vers la notion d'espèce naturelle, la nature essentialiste de leurs propos a provoqué des critiques variées. Parmi ces critiques, les difficultés auxquelles fait face toute tentative de rendre compte des catégories biologiques sur la base d'un essentialisme (même dénué de sa dimension téléologique) ont vite été soulignées par la communauté des philosophes de la biologie (Boyd 1991; Dupré 1993; Hull 1965a, 1965b; Sober 1980). Il est donc peu surprenant que les principales approches en compétition avec l'essentialisme aient été formulées par des membres de cette communauté. Je pense ici à Richard Boyd et à John Dupré qui ont développé les approches des groupes homéostatiques de propriétés (GHP) et du réalisme dévergondé (*promiscuous realism*)¹⁹, respectivement. Ces deux auteurs ont eu une influence marquée sur la tradition des espèces naturelles : leurs apports conjugués ont contribué à mettre de l'avant le pluralisme au sein de ce complexe discursif.

Dans ce qui suit, je présente donc leurs réflexions entourant la notion d'espèce naturelle afin de mettre en exergue la teneur spécifique de ce pluralisme avec lequel les théoriciens et théoriciennes doivent nécessairement négocier depuis qu'il est sur la table. J'en profite du même coup pour illustrer le fait que ces approches aussi, malgré les différences marquées entre elles et l'essentialisme, réitèrent l'arche des espèces naturelles.

3.3.2.1 Boyd, la thèse de l'accommodement et les groupes homéostatiques de propriétés

L'approche GHP des espèces naturelles a connu un succès indéniable dans les décennies qui ont suivi sa formulation par Boyd (1989, 1991, 1997, 1999a, 1999b). Ereshefsky et Reydon

¹⁹ Je dois l'utilisation du terme « dévergondé » pour traduire le terme « promiscuous » à Antoine Corriveau-Dussault. Je l'en remercie.

résumé comme suit l'importance de cette théorie (en introduction d'un texte dans lequel ils la critiquent) :

Richard Boyd's (1990, 1991, 1999a, b, 2000, 2003a, b) Homeostatic Property Cluster (HPC) Theory is becoming the received view of natural kinds in the philosophy of science. According to Ian Hacking, HPC Theory 'is the best recent contribution to the doctrine of natural kinds' (1991b, p. 149), and Samuels and Ferreira write that 'philosophers of science have, in recent years, reached a consensus—or as close to consensus as philosophers ever get—according to which natural kinds are Homeostatic Property Clusters' (2010, p. 222). Many philosophers of biology and the social sciences have adopted HPC Theory, including Griffiths (1999), Wilson (1999), Wilson et al. (2007), Brigandt (2009), and Mallon (2007). HPC Theory has even made its way into biology (for example, Wagner 2001; Keller et al. 2003; Rieppel 2005[b]). (Ereshefsky et Reydon 2015, 969-970)

Dans le cadre de mon projet, la thèse de Boyd revêt une importance particulière puisqu'il a été défendu à quelques reprises que les homologues (en plus des espèces biologiques) puissent être décrites comme des espèces naturelles dans un cadre GHP (Brigandt 2007; Assis et Brigandt 2009; Keller, Boyd et Wheeler 2003; Rieppel 2015; Wagner 1996), de sorte que les espèces naturelles GHP sont considérées comme étant importantes pour l'analyse de la phylogénétique. Dans le cadre de l'argument éliminativiste détaillé au chapitre 4, j'appréhende cet enjeu de front. Dans cette section-ci, je me concentre sur la présentation des éléments principaux de l'approche GHP.

La théorie GHP de Boyd se veut une solution de remplacement à l'essentialisme. Une espèce GHP est caractérisée par un groupe de propriétés (*property cluster*) qui tendent à être réalisées ensemble, dans la nature, parce qu'un mécanisme homéostatique assure cette réalisation commune. Selon cette approche, aucune propriété n'est nécessaire ou suffisante à l'appartenance d'un individu à l'espèce naturelle en question; c'est plutôt le mécanisme homéostatique qui permet d'individualiser les espèces naturelles. Il est important de préciser que ces mécanismes peuvent être extérieurs aux représentants de la catégorie. Par exemple, il peut s'agir de pressions environnementales qui tendent à engendrer la cooccurrence de propriétés au sein d'entités données.

L'espèce biologique est un exemple paradigmatique d'espèce GHP : la similarité observée entre les membres de l'espèce est expliquée par un ensemble de mécanismes causaux sous-

jacents (la fluidité génétique au sein des populations, l'appartenance des individus de l'espèce à des niches écologiques similaires ou tout simplement la présence d'hérédité) qui assurent l'homéostasie recherchée. Une conséquence de cette remarque est que les espèces GHP peuvent être à la fois des espèces naturelles et des entités historiques (les membres d'une espèce biologique étant liés par leur appartenance à une lignée au sein de laquelle les processus héréditaires variés se produisent, ils forment une entité qu'on peut isoler dans le temps et l'espace; Keller, Boyd et Wheeler 2003). Nonobstant cette possibilité d'un statut ontologique hybride, la correspondance des espèces GHP à des entités historiques demeure accessoire. Magnus (2014b), par exemple, suggère que les molécules d'H₂O sont au cœur de nexus causaux qui expliquent la stabilité de leurs propriétés sans que toutes les molécules d'H₂O soient liées par une histoire commune pour autant.

Ainsi, la théorie GHP, au même titre que l'essentialisme, est une description d'un fait ontologique relatif à des catégories scientifiques. Boyd est cependant explicite quant à la teneur pluraliste de sa conception des espèces naturelles : « I identify *a* class of natural kinds, properties and relations whose definitions are provided not by any set of necessary and sufficient conditions, but instead by a 'homeostatically' sustained clustering of those properties or relations. » (Boyd 1999a, 141, italiques ajoutés).

Magnus, dans une analyse éclairante des travaux de Boyd, insiste sur ce point. L'approche GHP est une manière de détailler l'ontologie de *certaines* des catégories utiles aux sciences, des catégories efficaces pour structurer notre rapport au réel et que Boyd considère comme des espèces naturelles. Dans la citation suivante, l'acronyme HPC renvoie aux groupes homéostatiques de propriétés et l'acronyme NK renvoie aux espèces naturelles : « So, the best understanding of HPCs withdraws from NK = HPC in several respects : HPCs provide *an* answer only to [...] the question of what it is in the world that holds natural kinds together. It turns out that many important natural kinds, although not all of them, are HPCs. » (Magnus 2014b, 476, italiques ajoutés). Ainsi, le pluralisme de Boyd vise le niveau ontologique. Au contraire, cet auteur influent adopte une posture moniste sur le plan épistémique : toutes les espèces naturelles ont en commun d'être utiles pour les explications et les inférences inductives réalisées en sciences avec succès (*successful inductive and explanatory practice in the sciences*).

Ce point est important parce qu'une critique souvent réitérée à l'égard de la théorie de Boyd est le fait qu'elle exclut certaines catégories scientifiques (Ereshefsky et Reydon 2015; Khalidi 2018). Cette critique est, à la lumière du pluralisme de Boyd, inopportune puisque son approche vise à décrire ontologiquement seulement un sous-ensemble des espèces naturelles. L'idée reçue selon laquelle la théorie GHP serait une théorie générale visant à couvrir l'ensemble des espèces naturelles est tout simplement erronée (du moins en se fiant à la version initiale de cette théorie qui a par la suite été interprétée de manière moniste par nombre d'auteurs et d'autrices). Bien au contraire, l'idée qui guide le développement de la théorie GHP est celle selon laquelle certaines catégories scientifiques considérées comme des espèces naturelles sont exclues par un essentialisme strict; il faut donc avoir une théorie de rechange pour rendre compte des espèces naturelles qui seraient autrement exclues, sans pour autant abandonner l'essentialisme utile.

Un autre point important lié à l'approche GHP est qu'elle s'inscrit au sein d'un paysage associé au réalisme scientifique. D'après cette posture, des divisions existent bien dans le monde naturel et les pratiques scientifiques font office de guide pour les traquer adéquatement. Conséquemment, les catégories scientifiques seraient celles desquelles les théories des espèces naturelles devraient rendre compte. Cela étant dit, le réalisme de Boyd doit être compris à la lumière de ce qui est appelé la thèse de l'accommodement (*accommodation thesis*).

La thèse de l'accommodement²⁰ est l'affirmation selon laquelle les catégories scientifiques tirent leur « réalité » non pas du fait d'une correspondance avec le monde tel qu'il existe indépendamment de nos tentatives d'en rendre compte, mais bien de leur capacité à nous faire interagir efficacement (en termes inférentiels) avec le monde empirique :

What the accommodation thesis indicates is that the metaphysical achievement that the deployment of kind terms and concepts may or may not represent is the

²⁰ Je base mon interprétation de la thèse de l'accommodement principalement sur un texte de Boyd (1999a) et sur une partie des écrits secondaires (par exemple, Ereshefsky et Reydon 2015; Franklin-Hall 2015; Magnus 2014a, 2014b). Une autre lecture de la chose est possible sur la base d'un plus vieux texte (Boyd 1991) qui réconcilie l'idée d'accommodement avec l'espoir de découper le monde en fonction de ses articulations naturelles. Cette seconde interprétation, qui associe à l'approche de Boyd une posture réaliste plus forte (Slater 2015), est moins présente dans le corpus étudié.

accommodation of inferential practices to relevant causal structures, so the 'reality' of a kind consists in the contribution that reference to it makes to such accommodation. (Boyd 1999a, 159)

La valeur métaphysique des espèces naturelles est, dans cette perspective, dérivée d'un fait épistémique à leur sujet : leur capacité à retracer des structures causales d'intérêt.

Cette posture n'en demeure pas moins réaliste à l'égard des catégories scientifiques : celles-ci nous permettent d'interagir efficacement avec le monde parce qu'elles se rapportent à certains aspects réels de ce monde (les espèces GHP, par exemple, renvoient à de réels phénomènes causaux qui expliquent la cooccurrence régulière de certaines propriétés). En revanche, la réalité des catégories scientifiques n'est reconnue que grâce à leur efficacité au sein d'une matrice disciplinaire (*disciplinary matrix*). De ce fait, elles ne traduisent qu'un aspect spécifique de la structure du monde; elles sont incapables de fournir une description de *la* structure du monde. C'est donc dire que le réalisme de Boyd accorde une place importante au contexte épistémique. Une espèce est dite naturelle seulement à la lumière de son utilité, elle-même entièrement dépendante de son imbrication dans un système de pensée relatif à une discipline scientifique donnée : « The naturalness of a natural kind is a matter of the contribution that reference to it makes to the satisfaction of the accommodation demands of a disciplinary matrix, in the context of a system of a compositional linguistic resources for the representation of phenomena » (Boyd 1999a, 158). Différentes matrices disciplinaires peuvent ainsi mobiliser des espèces naturelles différentes, tout aussi réelles les unes que les autres.

L'acception que Boyd lie à l'étiquette *réaliste* est le fruit d'une déflation telle que Boyd suggère ultimement de penser le réalisme exclusivement par l'entremise de l'accommodement des pratiques scientifiques au domaine empirique :

When we ask about the 'reality' of a kind or of the members of a family of kinds – or when we address the question of 'realism about' them – what we are addressing is the question of what contribution, if any, reference to the kind or kinds in question makes to the ways in which the classificatory and inferential practices in which they are implicated contribute to the satisfaction of the accommodation demands of the relevant disciplinary matrix. (Boyd 1999a, 159)

S'il y a une opposition forte entre les perspectives essentialistes et celle de Boyd (qui ne s'oppose aucunement à l'idée que certaines espèces puissent être analysées par l'entremise de

l'essentialisme), elle doit être comprise par l'entremise de cette thèse de l'accommodement et de ses conséquences. Pour l'essentialisme, l'utilité épistémique et la valeur métaphysique des espèces naturelles sont dérivées d'un fait ontologique, soit l'essence. Pour Boyd et une partie significative des philosophes ayant écrit sur les espèces naturelles dans une perspective non essentialiste depuis le début du 3^e millénaire de notre calendrier chrétien, la métaphysique et l'ontologie des catégories scientifiques sont le prolongement de constats épistémiques.

En bref, l'influence de Boyd sur les théories contemporaines des espèces naturelles emprunte trois voies distinctes, mais complémentaires. Premièrement, l'approche GHP permet de relâcher l'injonction essentialiste à associer les espèces naturelles avec des traits nécessaires et suffisants afin d'attirer l'attention des théoriciens et théoriciennes sur les groupes de propriétés qui tendent à être réalisées conjointement. Cet accent mis sur les groupes de propriétés est de nos jours très fréquent, même au-delà des théories GHP (par exemple, Khalidi 2018; Martínez 2020; O'Connor 2017; Slater 2015). Deuxièmement, Boyd suggère que des mécanismes causaux, parfois extrinsèques aux représentants d'une espèce naturelle, peuvent expliquer la stabilité de groupes de propriétés. Depuis, même les philosophes essentialistes (par exemple, Griffiths 1999) ont ajusté le tir pour intérioriser les constats de Boyd. D'autres philosophes, répudiant quant à eux l'essentialisme, se sont inspirés de l'accent mis par Boyd sur la notion de causalité (par exemple, Khalidi 2013). Finalement, la thèse de l'accommodement invite à penser les espèces naturelles comme étant des notions utiles plutôt que des concepts qui découpent le monde en fonction de ses articulations, tout en conservant l'intuition réaliste inhérente à la tradition des espèces naturelle. Cette utilité épistémique traduit en effet un statut métaphysique qui distingue certaines catégories des autres, les espèces naturelles des non naturelles, et qui invite une pluralité d'explications ontologiques.

3.3.2.2 Dupré et le réalisme dévergondé

La posture de Dupré, qu'il appelle lui-même le réalisme dévergondé (*promiscuous realism*), au sujet des espèces naturelles repose sur deux idées : premièrement, différentes manières de classer non concordantes peuvent être aussi légitimes les unes que les autres; deuxièmement, ces différentes classifications représentent tout autant d'aspects du monde tel qu'il existe indépendamment de nous, qu'elles émergent de pratiques scientifiques ou non. L'idée

sous-jacente est éminemment pragmatiste : tout schème classificatoire vise à répondre à des problèmes précis et circonscrit les catégories les plus aptes à contribuer à la résolution de ces problèmes (Dupré 1981, 1993). Ce constat vaut tout autant pour la phylogénéticienne qui classifie le vivant sur la base de liens de parenté que pour le jardinier qui classifie le vivant sur la base de qualités esthétiques, pratiques, économiques, etc. En effet, le monde est hautement complexe, de sorte que plusieurs manières de le décrire, de le découper et de discriminer ses éléments constitutifs peuvent être valides.

Dupré illustre sa vision par l'entremise de la catégorie des lys (*lilies*), qui réfère à une classe de plantes. Dans le langage courant (ou celui des jardiniers et des jardinières), le terme *lys* correspond principalement à un type de fleur. Or, ce type de fleurs regroupe des espèces qui, d'un point de vue phylogénétique, appartiennent à différents genres. À l'inverse, certaines espèces associées, en phylogénétique, à la famille des liliacées (l'équivalent phylogénétique de la catégorie *lys*) sont exclues de la catégorie *lys* du sens commun. Les tulipes ou les oignons, par exemple, ne sont pas reconnus par les jardiniers ou jardinières comme étant des lys. Il y a donc deux façons de penser la catégorie *lys* et toutes deux sont légitimes. Surtout, insiste Dupré, les deux catégories *lys* sont des espèces naturelles puisqu'elles sont construites sur la base d'aspects réels du monde. La validité d'une catégorie est, dans cette optique, tributaire du contexte dans lequel elle est mobilisée (le lys du jardinier étant efficace lorsque mobilisé dans le contexte jardinier).

La critique principale qui a été formulée à l'égard de ce réalisme dévergondé consiste à signaler qu'il rend triviale la catégorie d'espèce naturelle (Bird 2018; Wilson 1996). Comme mentionné précédemment, l'idée selon laquelle il y a une différence ontologique fondamentale entre les espèces naturelles et les autres catégories est solidement ancrée dans la tradition. Voici quelques formulations qui témoignent que cette croyance, modelée aux besoins de chacun et de chacune, est bien réactualisée par les itérations récentes peuplant le complexe discursif relatif aux espèces naturelles :

The problem with this simple association of natural kind classification with inductive success is that it promotes an overly liberal conception of natural kinds. Many successful inductions concern natural *relations* between objects (such as the

separation of objects and the forces they exert on one another) but entering into a natural relation typically does not form any basis for grouping entities into a natural kind. (Bird 2018, 1401)

Whether or not pluralism turns out to be correct, classification is not supposed to be 'anything goes,' with every conceivable grouping tracking a so-called *natural kind*. (Franklin-Hall 2015, 926)

On the other hand, we do not want to offer an account of kinds that merely says that any classification offered by a scientist picks out natural kinds. Consequently, we introduce the idea of a classificatory program and offer criteria for judging whether a classificatory program identifies natural kinds. The account of kinds we propose charts a middle course between HPC Theory's overly restrictive account of kinds and promiscuous accounts such as Dupré's (1993). (Ereshefsky et Reydon 2015, 970)

I contend that natural kinds are thought to provide an accurate understanding of the contents of the world conceived of as natural in at least three different ways: (1) that there are natural groupings, classifications, or methodologies that exist; (2) that natural kinds pick out these natural groupings; and (3) that theories of natural kinds specify the naturalness of kinds in their descriptions of the contents of the world. (Kendig 2016, 2)

C'est donc dire que l'ensemble du complexe discursif²¹ semble s'inscrire en opposition à la libéralisation du statut d'espèce naturelle suggérée par Dupré. Selon les auteurs et autrices cités, certaines catégories (scientifiques ou non) sont naturelles, d'autres (scientifiques ou non) ne le sont pas. Cela explique la charge qui a été portée à l'égard du réalisme dévergondé de Dupré, trop libéral au goût de certains et de certaines.

Cette critique fait pourtant abstraction d'un élément important de la théorie de Dupré, un élément pragmatiste s'il en est un, mais qui circonscrit aussi sa libéralisation de l'expression « espèce naturelle. » Selon lui, dans une perspective qui préfigure la thèse de l'accommodement de Boyd (telle que formulée en 1999a), les catégories qui méritent le statut d'espèce naturelle sont certes multiples et non concordantes, mais elles doivent minimalement faciliter le rapport de la cognition humaine au monde qui nous entoure. Les catégories scientifiques remplissent facilement cette condition, de plusieurs et de diverses façons (Ereshefsky et Reydon 2015), mais les catégories non scientifiques aussi, argumente Dupré. Que la catégorie « arbre » soit

²¹ Je suis conscient du fait que mon échantillon est limité. Or, la croyance selon laquelle les espèces naturelles sont différentes des espèces non naturelles est implicite dans l'expression en question et suffisamment fondamentale pour justifier la généralisation ici suggérée. L'omniprésence de l'arche des espèces naturelles renforce ce constat.

monophylétique ou non, elle facilite le rapport de bien des personnes à leur environnement en identifiant une catégorie utile d'un point de vue économique, communicationnel, relative au paysagement, etc. (Dupré 1981). Cette catégorie se distingue d'autres qui seraient, par exemple, formées complètement aléatoirement et qui seraient conséquemment nuisibles d'un point de vue épistémique.

Boyd a analysé la posture de Dupré d'une façon qui concorde avec mon commentaire. Selon lui, la différence entre l'approche dévergondée de Dupré et une vision plus stricte de ce qu'est une espèce naturelle relève du placement de la frontière entre espèces naturelles et non naturelles. Cette différence obscurcit toutefois un accord plus important sur ce qu'est une espèce naturelle, c'est-à-dire une entité utile pour structurer notre rapport au monde. Boyd lie directement la posture de Dupré avec sa thèse de l'accommodement en référant à l'exemple de la catégorie « lys » du sens commun mentionné précédemment : « This example illustrates an important fact: even the affairs of everyday life require accommodation between conceptual/classificatory resources and causal structures, so everyday kinds are usually natural kinds in the sense defined by the accommodation thesis. » (Boyd 1999a, 161) Boyd note tout de même une différence entre les catégories scientifiques et celles qui ne le sont pas, mais il s'agit d'une distinction de degré et non de nature : « I suspect that there is something like a continuum in degree of theoretical or integrative commitment between everyday accommodation-serving kinds and scientific natural kinds. » (Boyd 1999a, 161). Un peu plus loin, il ajoute la chose suivante : « [...] ordinary kinds and scientific natural kinds lie along a continuum. They do so precisely because they are all kinds of natural kinds – that is, resources for achieving accommodation » (Boyd 1999a, 162). Les approches de Boyd et de Dupré convergent donc indéniablement sur ce point : les espèces naturelles ont un statut épistémique privilégié lié au fait qu'elles rendent possible le contact efficace avec le monde qui nous entoure; ce statut épistémique fait des espèces naturelles des catégories métaphysiquement dignes de mention.

3.3.2.3 Les théories dites pluralistes, en bref

Les théories pluralistes à l'égard des espèces naturelles, autant celle de Boyd que celle de Dupré, le sont moins qu'on le considère généralement. *Elles sont pluralistes quant à la description ontologique des espèces naturelles ou de leurs représentants, mais monistes à l'égard du statut*

épistémique privilégié des espèces naturelles (à partir duquel on peut déduire le fait métaphysique de leur naturalité et justifier le réalisme à leur égard). L'innovation principale de ces deux auteurs est d'avoir clairement séparé la question épistémique relative aux espèces naturelles des descriptions ontologiques à leur égard²². Il y a ainsi une pluralité de descriptions ontologiques qui peuvent correspondre aux espèces naturelles, ces catégories qui facilitent notre contact cognitif avec le monde qui nous entoure. La pluralité ontologique est doublée d'une pluralité de systèmes classificatoires, la validité de chacun étant relative au système épistémique au sein duquel il émerge et est utilisé.

Les trois croyances fondamentales constituant l'arche de la tradition des espèces naturelles sont présentes chez Boyd et Dupré : 1) les espèces naturelles se distinguent métaphysiquement d'autres catégories (même si l'ontologie des espèces naturelles est plurielle, elle traduit des aspects réels du monde, ce qui témoigne de l'adhésion des deux auteurs au réalisme); 2) les espèces naturelles se distinguent aussi par leur statut épistémique (elles favorisent notre contact épistémique efficace avec le monde); 3) avoir *des* théories qui permettent d'appréhender diverses dimensions des espèces naturelles est donc primordial pour comprendre les pratiques classificatoires en sciences.

3.3.3 Les approches non pluralistes, non essentialistes

En utilisant l'appellation « non pluralistes, non essentialistes », je réfère à des approches récentes qui visent à fournir une description unitaire des espèces naturelles sans réduire celles-ci à des catégories essentialistes. Il s'agit donc de répondre aux travaux de Dupré et de Boyd sans retomber dans les écueils de l'essentialisme. Il va sans dire que ce corpus est éminemment hétérogène et vaste (Bird 2009; Craver 2009; Ereshefsky et Reydon 2015; Ereshefsky 2018; Franklin-Hall 2015; Hazelwood 2017; Khalidi 2013; Magnus 2015a; Martínez 2020; Martinez 2017; Slater 2015; Spencer 2016). Il m'est donc impossible de le survoler de manière exhaustive. Pour pallier cette limitation, j'ai choisi deux approches distinctes, soit celles de Khalidi et de Slater, qui

²² Magnus (2014b, 2015a, 2015b) argumente qu'on peut lire Mill de manière à lui attribuer cette distinction. Nonobstant la plausibilité de cette interprétation des travaux de Mill, il convient minimalement d'attribuer la *popularisation* de la distinction en question à Boyd et à Dupré.

présentent un rapport très différent à l'unicité de la notion « espèce naturelle. » Alors que Khalidi perçoit cette unicité sur le plan ontologique, Slater insiste sur l'importance des caractéristiques épistémiques des espèces naturelles. En ce sens, leurs travaux sont représentatifs des principales tendances dans le corpus contemporain relatif aux espèces naturelles. Dans ce qui suit, je démontre que malgré les différences apparentes entre ces deux approches, celles-ci partagent une adhésion indéniable à l'arche des espèces naturelles. Les trois croyances concernées sont répliquées ici avec fidélité, de sorte qu'il est possible de tracer une ligne directe entre les travaux des philosophes anglais du 19^e siècle et les écrits contemporains, en passant par les textes de la deuxième moitié du 20^e siècle.

3.3.3.1 Khalidi et les espèces naturelles en tant que nœuds au sein de réseaux causaux

Le fait que Khalidi adhère à l'arche des espèces naturelles est parfaitement illustré par la citation suivante : « We infer what natural kinds exist and discern their commonalities after determining which categories play a central role in our knowledge-gathering enterprises. » (Khalidi 2018, 1385-1386). D'abord, la philosophie des sciences doit isoler les catégories détenant un rôle important au sein des pratiques scientifiques. De cette observation, il est possible d'inférer l'existence d'espèces naturelles et de discerner les points communs à l'ensemble de ces espèces. Une théorie des espèces naturelles revient, dans cette perspective, à décrire ces points communs et à rendre compte philosophiquement.

Selon la théorie de Khalidi, les espèces naturelles renvoient à des groupes de propriétés qui tendent à être coréalisées en raison de l'existence d'une structure causale sous-jacente. Plus spécifiquement, la structure causale réfère à la relation de causalité entre des propriétés fondamentales (*core properties*) et des propriétés dérivatives (*derivative properties*). La causalité émane des propriétés fondamentales, alors que ses effets sont ressentis sur le plan des propriétés dérivatives. Une espèce naturelle serait ainsi individualisée par l'architecture causale caractéristique d'un groupe de propriétés, une architecture qui se répète d'une occurrence à l'autre. Un organisme pourrait ainsi être décrit sur la base de propriétés fondamentales (gènes, génome, modules développementaux, etc.) qui expliquent causalement l'émergence de traits dérivatifs (taille, couleur des yeux, etc.). Comme cet exemple le suggère, les propriétés fondamentales peuvent être extrinsèques. En effet, on pourrait considérer que la niche

écologique dans laquelle se développe un organisme fait partie des propriétés fondamentales de cet organisme. La théorie de Khalidi se distingue ainsi de l'essentialisme par sa façon de caractériser l'ontologie des catégories et par la diversité des structures causales associée aux espèces naturelles : « on the current proposal, the causal properties associated with natural kinds can be extrinsic or functional, and there may be no set of properties that are singly necessary and jointly sufficient for membership in the kind. » (Khalidi 2018, 1385).

Cette théorie est moniste sur le plan ontologique en ce sens qu'elle considère que la description ontologique des espèces naturelles axée sur les réseaux de causalité permet de rendre compte de toutes les espèces naturelles et que seulement les catégories qui correspondent à cette description devraient se voir attribuer le statut d'espèce naturelle. L'approche de Khalidi se distingue donc de celles de Boyd et de Dupré sur ce point. Une autre différence relève de leurs versions respectives du réalisme. Pour Khalidi, l'existence d'espèces naturelles peut être inférée sur la base du travail des scientifiques, ce qui suggère, selon lui, qu'elles existent indépendamment de ce travail. Pour Boyd et Dupré, comme souligné précédemment, les espèces naturelles n'existent qu'en tant que constructions théoriques issues de matrices disciplinaires et facilitant notre rapport au monde (bien qu'elles témoignent d'aspects réels du monde). Khalidi pousse le réalisme plus loin en suggérant l'existence d'espèces naturelles au-delà des théories qui mobilisent les concepts théoriques correspondants (à la manière de Locke et de l'arrimage entre essences réelles et nominales).

L'approche de Khalidi se distingue aussi de l'approche GHP par son rapport à la causalité. Selon lui, il faut relâcher l'approche de Boyd, de sorte à pouvoir inclure une plus grande diversité de structures causales. Limiter les espèces naturelles aux groupes de propriétés dont la cooccurrence est expliquée par *un* mécanisme causal et *homéostatique* est trop restrictif. Pour Khalidi, même certaines caractéristiques des éléments chimiques doivent être expliquées par une structure causale complexe irréductible à un mécanisme unique. De surcroît, si l'homéostasie doit être comprise comme référant à la stabilité d'une occurrence d'un groupe de propriétés²³, alors

²³ L'homéostasie impliquée dans l'approche GHP est plus souvent interprétée comme référant au fait que les propriétés tendent à être réalisées communément et non au fait qu'une fois réalisées, elles demeurent stables,

cette condition vient exclure la possibilité pour les individus associés à une espèce naturelle de changer au travers du temps. Il s'agirait conséquemment d'une condition trop contraignante, ce qui justifie le relâchement définitionnel prescrit par Khalidi : « Boyd's account has to be loosened in such a way as to retain the emphasis on causality without the mechanism or the homeostasis. » (Khalidi 2018, 1386).

En bref, Khalidi considère que tout processus causal qui implique des liens entre propriétés génère de fait des nœuds causaux (*causal nodes*) qui peuvent eux-mêmes être impliqués au sein d'autres processus causaux (un échantillon d'or détient sa propre structure causale qui, en tant que nœud stable, peut être impliquée au sein d'autres processus causaux, en tant que valeur refuge, par exemple). Le rôle crucial de ces nexus causaux explique pourquoi et comment les concepts théoriques correspondants contribuent aux succès épistémiques des sciences.

Il est donc important de noter que l'approche de Khalidi est construite sur la base d'une démarcation entre corrélation et causalité, ce qui revient dans ce cas-ci à démarquer les espèces naturelles et non naturelles. En effet, Khalidi insiste sur le fait qu'il est parfois trompeur d'attribuer à un concept ou à une catégorie le statut d'espèce naturelle seulement sur la base de sa projectibilité, c'est-à-dire de sa capacité à soutenir des inférences inductives. Cela peut être trompeur puisque l'espèce naturelle ainsi identifiée pourrait référer à la cooccurrence de deux effets d'une cause sans rendre honneur à celle-ci; elle pourrait aussi impliquer l'inversion d'une chaîne causale ou encore lier des propriétés dont la cooccurrence n'est que le fruit du hasard. De ces potentiels écueils, écrit Khalidi, deux problèmes épistémologiques peuvent émerger : d'abord, cela peut mener à la mauvaise compréhension de ce qu'est une espèce naturelle (on pourrait croire qu'il s'agit de catégories identifiées sur la base de corrélations plutôt que sur la base de causalité); ensuite, cela peut mener à de faux positifs, à l'identification d'espèces naturelles qui n'en sont point.

En résumé, l'approche de Khalidi est ancrée dans la distinction entre espèces naturelles, soit les espèces qui expliquent les succès épistémiques des pratiques scientifiques, et les espèces

l'approche de Boyd ayant été développé pour rendre compte de groupes d'entités changeantes, notamment les espèces biologiques (Griffiths 2007; Wilson, Barker et Brigandt 2007). Khalidi en offre cependant une autre lecture.

non naturelles, soit des catégories qui sont privées d'une structure causale adéquate. La structure logique de base de la tradition des espèces naturelles est encore une fois réitérée : le statut métaphysique est lié à la spécificité épistémique des espèces naturelles, le premier expliquant la seconde et le tout justifiant la nécessité pour la philosophie des sciences de se munir d'une théorie des espèces naturelles.

3.3.3.2 Slater et les groupes stables de propriétés

La théorie des groupes stables de propriétés prend très au sérieux la thèse de l'accommodement de Boyd (Slater 2015 explicite cette approche, mais voir aussi Slater 2013 et 2018). L'idée centrale de Slater est le fait que toute métaphysique entourant les espèces naturelles ne peut être tributaire que de leur valeur épistémique : « Modest realists, for instance, can interpret the 'projectibility' of certain categories as amounting to a metaphysical fact about those categories and a epistemological matter concerning our recognition of this fact (presumably in the context of much background knowledge). » (Slater 2015, 385). Il s'agit là d'une reformulation de la thèse de l'accommodement que Slater utilise pour articuler l'ensemble de sa théorie²⁴. Comme pour la théorie de Dupré, le monisme est ici camouflé par un revêtement pluraliste que la théorie (et ses critiques) met de l'avant : Slater reconnaît que différents fondements ontologiques (mécanismes homéostatiques, essences, etc.) sont attribuables à différentes espèces naturelles, mais il insiste sur le fait qu'une description unitaire des espèces naturelles est possible sur le plan épistémique.

À la manière de Dupré et de Boyd, Slater renverse la logique associée à l'essentialisme ou à l'approche de Khalidi : il est trompeur de présumer que la grande variété des catégories scientifiquement utiles détient une structure ontologique fondamentale commune (une essence ou structure causale); cependant, le fait même de leur utilité peut être compris comme une caractéristique unificatrice décrivant toutes les catégories détenant le statut d'espèce naturelle.

²⁴ Slater considère que la thèse de l'accommodement implique la possibilité de découper le monde en fonction de ses articulations naturelles (*to carve the world at its joints*), son interprétation étant basée sur un vieux texte de Boyd (1991). La lecture plus commune de ladite thèse (Ereshefsky et Reydon 2015; Franklin-Hall 2015; Magnus 2014b), à laquelle j'adhère, concorde plutôt avec les textes tardifs de Boyd (1991, 2013). Selon cette seconde lecture, la posture de Slater apparaît effectivement comme une reformulation de la thèse de l'accommodement.

Bien que pluraliste à l'égard de l'ontologie, cette approche est effectivement moniste à l'égard des caractéristiques épistémiques des espèces naturelles : « my account does not advertise itself as an account of a kind of natural kind—it is, I think, an attractive candidate for a general natural kind concept, able to accommodate the diversity of natural kinds we find in the world. » (Slater 2015, 377).

Afin de rendre compte de cette diversité d'espèces naturelles qu'on retrouve dans le « monde », Slater offre la notion de stabilité de quasi-clique (*cliquish stability*) à penser dans une perspective pragmatiste sensible au contexte épistémique au sein duquel les espèces naturelles sont mobilisées. La stabilité de quasi-clique réfère à la projectibilité des espèces naturelles, c'est-à-dire à la cooccurrence régulière des propriétés qui leur sont associées et qui permet des inférences inductives au sujet de leurs représentants. Si l'espèce naturelle *E1* est caractérisée par les propriétés *P1*, *P2* et *P3*, alors on peut s'attendre à ce qu'un individu détenant *P1* détienne aussi *P2* et *P3*. La détention d'une seconde propriété renforce la possibilité que la troisième soit présente, etc. L'idée de stabilité de quasi-clique précise que cette cooccurrence implique seulement l'émergence régulière et conjointe des propriétés, sans que cette stabilité doive impliquer la pérennité de la relation.

Slater suggère aussi que la stabilité requise pour qu'un groupe de propriétés cooccurrentes puisse être associé à une espèce naturelle dépend du contexte épistémique en question. Par exemple, la stabilité requise pour faire d'une particule physique élémentaire une espèce naturelle est plus grande que celle requise pour faire d'une espèce biologique une espèce naturelle, mais cette distinction laisse l'unicité du concept d'espèce naturelle intacte. En effet, tant qu'une catégorie est suffisamment stable pour répondre aux besoins des chercheurs et chercheuses concerné-e-s, cette catégorie peut aspirer au statut d'espèce naturelle. À la manière de Boyd qui considère que l'accommodement des concepts théoriques aux phénomènes est relatif aux matrices disciplinaires, le seuil de stabilité mis de l'avant par Slater pour les espèces naturelles est relatif au contexte épistémique concerné.

Cela étant dit, comme chez Boyd et Dupré, la croyance selon laquelle les espèces naturelles sont des entités épistémiquement et métaphysiquement particulières est maintenue

et justifie le développement d'une théorie des espèces naturelles. L'arche des espèces naturelles est bien présente puisque Slater construit l'ensemble de sa théorie afin de rendre compte de la distinction entre les catégories qui méritent le statut d'espèce naturelle et celles qui en sont indignes. Dans la citation suivante, l'acronyme « SPC » réfère à la théorie des groupes stables de propriétés de Slater : « What the SPC account offers us is a flexible, high-level approach to understanding the various ways in which various categories (scientific or otherwise) can be regarded as *genuine features of the world* in organizing and facilitating our epistemic contact with the world. » (Slater 2015, 407, italiques ajoutés). Les espèces naturelles représentent des « caractéristiques authentiques du monde », de sorte qu'il est important de se doter d'une théorie philosophique des espèces naturelles pour comprendre leur imbrication au sein des pratiques classificatoires ainsi que leur portée empirique.

3.3.3.3 Les approches non pluralistes non essentialistes, en bref

Cette brève présentation de deux théories des espèces naturelles à la fois récentes et influentes (Brigandt 2020) témoigne de la pérennité de l'arche des espèces naturelles. La justification même pour le développement d'une théorie des espèces naturelles relève des deux piliers se renforçant mutuellement : certaines catégories scientifiques jouent un rôle important au sein des pratiques épistémiques et détiennent un statut métaphysique particulier. La seule chose qui sépare, à ce niveau d'abstraction, les théories de Khalidi et de Slater est la direction du lien unissant les deux piliers. Pour Khalidi, la spécificité ontologique des espèces naturelles explique leur utilité épistémique (bien qu'on ne puisse identifier les espèces naturelles seulement en se fiant aux succès des pratiques scientifiques), alors que pour Slater, le fait métaphysique est l'extension directe du fait épistémique. Nonobstant cette différence, les deux auteurs partagent ultimement la croyance fondamentale selon laquelle une théorie des espèces naturelles peut nous aider à comprendre les pratiques classificatoires.

Conjointement avec la première partie du chapitre, cette partie m'a permis de démontrer la stabilité de l'arche des espèces naturelles au sein du corpus étudié. Cette stabilité s'explique comme suit : une théorie philosophique de la classification qui rejette cet assemblage de croyances est tout simplement autre chose qu'une théorie des espèces naturelles. En effet, si le fait de répudier le statut métaphysique des espèces naturelles, leur statut épistémique et

l'injonction à les théoriser était cohérent avec la formulation d'une théorie des espèces naturelles, cela témoignerait de fait de la trivialité de la notion. J'en conclus, sur la base de l'étude d'un échantillon limité, mais significatif quant à la diversité des théories analysées, que l'arche des espèces naturelles est inhérente aux principales théories des espèces naturelles.

3.4 Conclusion

Dans ce chapitre, j'ai démontré que la tradition des espèces naturelles, en plus d'être unie par l'utilisation de l'expression en question (Conix et Chi 2020), s'est cristallisée autour d'un complexe de trois croyances que j'appelle l'arche des espèces naturelles. Cette arche est déclinée différemment par chacune des approches portant sur les espèces naturelles, mais son essence demeure intacte au sein de chaque itération. Cette arche est constituée de deux piliers : la croyance selon laquelle des catégories scientifiques ont un statut métaphysique particulier et celle selon laquelle les mêmes catégories scientifiques ont un statut épistémique privilégié. Les deux piliers sont habituellement renforcés par un lien logique qui les unit : soit la métaphysique explique le fait épistémique, soit le fait métaphysique est inféré sur la base du constat épistémologique au sujet des catégories scientifiques. Dans tous les cas, ces deux piliers justifient, au moins implicitement, la formulation d'une ou de plusieurs théories des espèces naturelles (l'arc de l'arche) :

An underlying assumption of this paper is that a proper philosophical account of natural kinds should help us understand the success of natural kind classifications in science. Like many endeavors, a study of natural kinds should study best practices, and arguably those best practices are classificatory practices in science. This idea is far from new. It is line with numerous theories of natural kinds, from Locke (1894) to Boyd (1999[a]). (Ereshefsky 2018, 845)

Cette citation d'Ereshefsky démontre que l'arche des espèces naturelles est plus qu'une structure interne aux quelques textes étudiés dans ce chapitre. Cette arche sous-tend, au sein de l'ensemble de la tradition des espèces naturelles, l'injonction à appréhender les pratiques classificatoires sur la base de l'étude des catégories produites, utilisées et candidates au statut d'espèce naturelle. Ainsi, des théories aussi distinctes que celles dites essentialistes, celles dites pluralistes, celles se voulant monistes et celles inscrites au sein des textes classiques de la tradition incorporent toutes l'arche des espèces naturelles, de sorte que ce courant de pensée

puisse effectivement être considéré unitaire en tant que posture épistémologique à l'égard des pratiques classificatoires.

Chapitre 4 – Éliminativisme et espèces naturelles

Un jour, c'était l'hiver, je suis tombée sur la glace et j'ai eu un œil au beurre noir. [...] Si j'avais été une femme blanche, j'aurais été [aux yeux des personnes qui remarquent l'œil au beurre noir] cette pauvre fille battue. Ou empotée. Ou violente. J'aurais été une parmi d'autres à subir une relation malsaine avec son amoureux et on m'aurait plaint pour ça. Personne n'aurait été associé à mon malheur. Il aurait été unique, individuel.

Mais je suis Innue.

Cette blessure sur mon visage n'était pas seulement la mienne, elle appartenait aussi à ma nation. Elle incarnait, et j'en étais aussi consciente que désolée, la blessure des Indiennes. Plutôt, celle de tous les préjugés vis-à-vis les femmes autochtones disparues et assassinées. Dans le meilleur des cas, mon œil confirmait que nous, les femmes innues, étions des victimes. Dans les pires jugements, nous étions violentes, agressives, jusqu'à nous battre dans les bars ou dans la rue. Sans savoir-vivre. Comme j'ai eu honte de moi, Shuni.

Rarement, les gens me perçoivent comme un individu unique. Dans un groupe, on ne m'appelle pas par mon nom. On dira l'Indienne, l'Innue, l'Autochtone. Si je tombe, c'est tous les autres qui tombent avec moi. Si je me tiens debout, ils sauront que nous sommes résistants.

- Naomi Fontaine, dans *Shuni*

4.1 Introduction

Le chapitre précédent m'a permis de démontrer l'unité de la tradition des espèces naturelles en philosophie des sciences. Celle-ci est guidée par un triumvirat de croyances qui lui sont inhérentes : 1) certaines catégories scientifiques sont métaphysiquement distinctes des autres catégories (scientifiques ou autres); 2) ces mêmes catégories jouent un rôle privilégié au sein des activités épistémiques des communautés scientifiques; 3) il est important d'avoir une ou des théories des espèces naturelles afin de rendre compte des pratiques classificatoires en sciences. J'ai démontré que ces trois croyances forment une arche et que cette arche est associée à une forme de naturalisme. Ce naturalisme, traduit par le lien entre les deux piliers de l'arche, suggère que les catégories scientifiques sont, du fait de leur utilité épistémique, à considérer comme les meilleurs guides pour la métaphysique. Dans ce chapitre-ci, j'expose en détail ma posture éliminativiste à l'égard des espèces naturelles, qui remet en question l'utilité d'analyses épistémologiques des pratiques classificatoires qui seraient structurées sur le modèle de l'arche des espèces naturelles.

Avant de procéder, il convient de préciser ce qu'est l'éliminativisme, et plus spécifiquement, ce qu'est l'éliminativisme à l'égard des espèces naturelles. L'éliminativisme est une posture qui suggère de reléguer aux oubliettes un concept donné. Loin d'être une posture générale, l'éliminativisme ne prend sens qu'en ciblant spécifiquement certaines notions, les motivations peuvent être différentes dans chaque cas. Par exemple, sur la base d'un pluralisme observé dans la pratique des biologistes, Ereshefsky (1992) a fameusement argumenté que de la notion d'espèce (*species*) devrait être abandonnée. Il est aussi possible de défendre un éliminativisme à l'égard de différentes catégories mentales (comme les états mentaux), et ce, pour différentes raisons (Ramsey 2021). En ce qui concerne le travail effectué dans cette thèse, la posture éliminativiste ne concerne que la notion d'espèce naturelle, même lorsque j'utilise le terme « éliminativisme » sans précision supplémentaire.

Pour bien comprendre l'éliminativisme à l'égard des espèces naturelles, il faut d'entrée de jeu saisir pourquoi la philosophie des sciences a jusqu'à présent jugé qu'il serait pertinent d'utiliser cette notion. Il y a minimalement deux raisons pour lesquelles l'épistémologie devrait s'intéresser aux espèces naturelles. Premièrement, on peut s'intéresser à la notion d'espèce naturelle parce que celle-ci est utile épistémologiquement parlant, c'est-à-dire qu'elle nous aide à comprendre les processus épistémiques étudiés. Cette première option est celle considérée dans la présente thèse. Ma posture éliminativiste vise donc à démontrer que l'utilité épistémologique du concept d'espèce naturelle est limitée et que de ce fait, il serait souhaitable d'en faire l'économie au profit d'une approche inspirée du cohérentisme historique. Mon approche se distingue donc d'éliminativistes articulés à l'égard de *certaines* espèces naturelles. Par exemple, l'éliminativiste à l'égard des espèces naturelles en psychologie (Churchland 1981; 1988) est beaucoup plus restreint que la posture que je défends.

Deuxièmement, il pourrait être important de s'intéresser aux espèces naturelles simplement parce que celles-ci existent. Il s'agirait alors d'un phénomène qui, comme bien d'autres phénomènes, mérite l'attention des chercheurs et chercheuses. Les espèces naturelles existent minimalement en tant que concept au sein de la culture intellectuelle occidentale. En ce sens, elles méritent d'être étudiées dans une perspective d'histoire des idées ou d'histoire de la

philosophie. Un argument éliminativiste à cet égard relèverait de la censure et serait nuisible plutôt que constructif.

Les espèces naturelles pourraient aussi exister en tant que phénomène dans la structure du monde. Au sein de la tradition des espèces naturelles, la posture réaliste à l'égard des espèces naturelles qui véhicule la croyance selon laquelle les espèces naturelles existent effectivement, minimalement en tant que divisions au sein de la structure du monde, est très fréquemment soutenue (Bird 2018; Crane 2003). Si cet aspect du débat n'est pas celui auquel je me frotte dans ce chapitre et cette thèse, je considère tout de même qu'un argument éliminativiste à cet égard est possible et légitime considérant les conséquences épistémiques du problème de la coordination (nous n'avons aucun accès à la structure du monde qui serait indépendant de nos modèles et de nos théories). Je garde mes intuitions à ce sujet pour structurer de futurs travaux et me concentre, dans ce chapitre-ci, sur l'aspect épistémologique de la chose.

Ce chapitre est composé de deux parties. Dans un premier temps, j'explore deux approches éliminativistes déjà présentes au sein du corpus de textes (Brigandt 2020; Ludwig 2018) et un argument contre l'éliminativisme (Conix et Chi 2021). En plus de compléter le survol du corpus associé aux espèces naturelles que j'ai amorcé au chapitre précédent, cela me permet de souligner les enjeux épistémologiques soulevés par l'éliminativisme à l'égard des espèces naturelles. De surcroît, l'analyse du texte de Conix et Chi me permet de poser les bases de mon argument éliminativiste, complémentaire à ceux de Brigandt et de Ludwig.

Dans un second temps, je présente trois arguments qui composent l'essentiel de ma posture éliminativiste. J'y mobilise la notion d'homologie comme point de référence puisque celle-ci a régulièrement été considérée comme étant candidate au statut d'espèce naturelle (dans Assis et Brigandt 2009 ou Rieppel 2015, par exemple). Le premier argument est une réponse directe à Brigandt (2009) qui affirme que la notion d'espèce naturelle GHP est épistémologiquement utile parce que ce fait métaphysique relatif aux espèces naturelles expliquerait leur utilité épistémique (réitération du lien entre les piliers de l'arche). De plus, Brigandt (2009) affirme que sans les concepts d'espèces naturelles, la connaissance scientifique ne pourrait pas être incarnée de façon efficace. Je formule mon argument contre celui de Brigandt

(2009) de sorte qu'il puisse éventuellement servir de réponse à d'autres théories des espèces naturelles et non seulement à la théorie GHP. En bref, cet argument stipule que l'étiquette *espèce naturelle* est attribuée *a posteriori*, après l'analyse épistémologique des pratiques classificatoires. Elle est donc superflue lors de ces analyses et ne prend sens que si l'objectif de la recherche consiste à identifier des espèces naturelles.

Mon deuxième argument vise à démontrer que les théories des espèces naturelles sont des mécanismes de légitimation des pratiques classificatoires scientifiques et qu'elles sont ainsi internes à l'objet d'étude de l'épistémologie de la classification. En ce sens, il serait peu judicieux de les utiliser pour analyser des pratiques classificatoires dont elles font partie. Mon troisième argument vise à démontrer que le fait de centrer les descriptions et analyses épistémologiques des pratiques classificatoires sur leur produit fini, soit les catégories scientifiques, mène à négliger la structure épistémique plus englobante qui devrait être l'objet d'étude premier de l'épistémologie de la classification. Le tout offre de bonnes raisons de considérer l'éliminativisme à l'égard des espèces naturelles comme une posture légitime en épistémologie de la classification, le fardeau de la preuve étant du même coup transféré aux philosophes qui estiment que la notion ciblée est importante pour l'analyse des pratiques classificatoires.

4.2 L'éliminativisme et les espèces naturelles

4.2.1 Deux arguments éliminativistes

Dans cette section, je présente deux arguments éliminativistes présents au sein du corpus portant sur les espèces naturelles. Le premier, celui de Ludwig (2018), s'inscrit en continuité avec celui de Hacking (2007a) et repose sur la pluralité des théories des espèces naturelles disponibles et sur l'utilité épistémologique de cette diversité. L'argument de Brigandt (2020) est quant à lui centré sur la « naturalité » des espèces naturelles et sur les problèmes épistémiques et épistémologiques que cette naturalité tend à faire émerger.

4.2.1.1 Un cadre d'analyse multidimensionnel pour appréhender les pratiques de classification non arbitraires

Dans un texte de 2007, Hacking (2007a) illustre la grande diversité des catégories scientifiques potentiellement titulaires du statut d'espèce naturelle et la tout aussi grande diversité des théories des espèces naturelles ayant été formulées dans l'espoir d'en rendre compte. Selon Hacking, cette hétérogénéité traduit un désordre au sein de la tradition des espèces naturelles qui ne peut aller qu'en s'intensifiant alors que le regard des philosophes se fait continuellement plus incisif et porte sur une diversité grandissante de pratiques classificatoires (Hacking 2007a). Les années ayant suivi cette publication ont donné raison à Hacking, alors que les théories des espèces naturelles se sont multipliées au sein de la communauté. Ce point témoigne aussi du fait que l'argument de Hacking est loin d'avoir conquis les masses (Ludwig 2018).

Dans son texte de 2018, Ludwig cherche à renforcer l'argument éliminativiste basé sur la diversité des théories. Le point central est le même que celui de Hacking : la diversité des théories des espèces naturelles et des catégories scientifiques candidates pour l'obtention de ce statut suggère que l'unité du concept « espèce naturelle » est douteuse et qu'il faudrait, en conséquence, l'abandonner. Pour Ludwig, cela implique à la fois un argument négatif contre l'idée d'une théorie des espèces naturelles unitaire et un argument positif mettant de l'avant un cadre de pensée multidimensionnel pour l'analyse de pratiques classificatoires non arbitraires.

Dans un premier temps, Ludwig démontre que le désir d'unifier l'usage du concept d'espèce naturelle implique une perte d'information. En effet, l'adoption d'une théorie unique s'accompagne de l'abandon de certaines théories des espèces naturelles puisqu'elles sont (au moins partiellement) incompatibles les unes avec les autres. Cet abandon est problématique puisque les théories abandonnées peuvent être utiles. Afin d'étoffer son argument, Ludwig mobilise un cas d'étude, soit l'analyse de la diversité des systèmes classificatoires que des êtres humains mobilisent pour décrire le monde vivant, une diversité culturelle étudiée par l'ethnobiologie. Ludwig présente alors une taxonomie des types de catégories qui sont mobilisées par ces différentes cultures : les catégories basées sur des usages spécifiques (la catégorie « champignons comestibles », par exemple), celles basées sur des usages généraux (la catégorie

« animaux volants », par exemple), les catégories convergentes et indépendantes de la cognition humaine (certaines catégories de vertébrés présentes au sein de différentes cultures, par exemple), les catégories convergentes et dépendantes de la cognition humaine (la catégorie « arbre », par exemple), des catégories dépendantes de la pratique (la catégorie « Chichlu'um », par exemple, qui réfère à un type de sol dans la culture maya), les catégories divergentes et dépendantes de l'environnement (la catégorie « Chichlu'um » peut encore servir d'exemple puisqu'elle est dépendante non seulement de certaines pratiques d'agriculture spécifiques, mais aussi des spécificités du sol du Yucatán), les catégories biosociales (certaines espèces de microbes identifiées sur la base de préoccupations sanitaires, par exemple).

Inutile ici de présenter le détail de chacun de ces types de catégories qui sont des outils spécifiques à l'ethnobiologie. Il s'agit simplement de saisir le rôle qu'ils jouent dans l'argument de Ludwig. Chacun de ces outils est important pour rendre compte des systèmes de classification du monde biologique mobilisés par une ou plusieurs cultures humaines et du rapport que ces systèmes entretiennent les uns avec les autres. Après avoir exploré quatre théories des espèces naturelles récentes et influentes (Franklin-Hall 2015; Khalidi 2013; Magnus 2014a, 2015a; Slater 2015), Ludwig constate qu'aucune d'entre elles n'est capable de rendre compte de l'ensemble des outils de l'ethnobiologie, mais que chacune facilite l'analyse épistémologique de certains d'entre eux. Il serait donc nuisible de se priver de certains de ces outils dans l'espoir d'acquérir une théorie unifiée des espèces naturelles.

L'argument de Ludwig est d'abord et avant tout un argument pluraliste, comparable à celui qu'Ereshefsky a formulé à l'égard du concept d'espèce en biologie (Ereshefsky 1992). Selon Ludwig, il y a différentes façons de rendre compte d'une diversité de phénomènes classificatoires et du fait qu'ils sont épistémiquement utiles (non arbitraires). Une fois ces différentes approches mobilisées, il n'y aurait aucun gain épistémologique à taire la précision des descriptions qui en résulte au profit d'un concept unitaire des espèces naturelles :

For example, kinds are often not arbitrary because they meet success and restriction clauses, constitute categorical bottlenecks, involve SPCs, or qualify as nodes in a causal net work. Once we understand all of this, however, there is no interesting epistemic or metaphysical work left for a general notion of natural kind. (Ludwig 2018, 47)

La richesse épistémologique du corpus associé aux espèces naturelles va de pair avec sa diversité, un fait irréconciliable avec l'injonction à identifier une théorie unique des espèces naturelles.

En s'appuyant sur ce constat, Ludwig propose, dans le second temps de son analyse, un cadre de pensée multidimensionnel pour l'analyse des pratiques classificatoires non arbitraires. Ce cadre mobilise la diversité des outils développés au sein des théories des espèces naturelles sans chercher à en prioriser une au détriment des autres. De plus, cette perspective permet d'appréhender, au sein d'une analyse épistémologique, des catégories éminemment non naturelles (des catégories socialement circonscrites) tout autant que des catégories ayant par le passé aspiré au statut d'espèce naturelle. Ce point est important puisque peu importe le statut métaphysique des catégories mobilisées, les pratiques classificatoires qui les sous-tendent peuvent être comparables et d'un intérêt tout aussi grand pour l'épistémologie.

L'argument de Ludwig est puissant, notamment parce qu'il démontre la possibilité de croiser les gains explicatifs de plusieurs théories habituellement jugées comme étant incompatibles. Pour ce faire, il faut cependant rejeter l'idée de l'unité des espèces naturelles inhérente à la tradition des espèces naturelles (que cette unité soit épistémique ou métaphysique). En ce sens, son texte déconstruit l'arche des espèces naturelles : il y a très certainement des catégories scientifiques avec un statut épistémique particulier, mais cela n'implique aucun statut métaphysique privilégié puisque certaines de ces catégories sont artificielles en tous points. En conséquence, la diversité des catégories et des manières d'en rendre compte remet en question la croyance selon laquelle il faut se munir de théories des espèces naturelles. D'un point de vue épistémologique, il faut plutôt se doter d'outils variés pour analyser des catégories non arbitraires en général.

Un autre point important à noter est une distinction inscrite implicitement au cœur du texte de Ludwig : il est possible de rejeter la notion d'espèce naturelle sans pour autant rejeter les fruits du travail qui a été accompli au sein de la tradition des espèces naturelles. Un argument éliminativiste suggère le rejet d'une notion et de son rôle structurant au sein de certaines postures philosophiques (ici associées à la tradition des espèces naturelles) et non le rejet de

l'ensemble du contenu du complexe discursif associé à cette notion. Cette précision est importante pour répondre à certaines critiques qui sont portées à l'égard des approches éliminativistes (par exemple, Conix et Chi 2021). J'y reviens plus loin.

4.2.1.2 La naturalité des espèces naturelles et l'influence des valeurs en sciences

Brigandt (2020) a récemment défendu une posture éliminativiste à l'égard des espèces naturelles qui s'articule elle aussi en deux temps. D'abord, il offre un cadre méthodologique pour appréhender les catégories en général et non seulement les espèces naturelles. Ce cadre invite à penser les catégories scientifiques à l'aune des objectifs (*aims*) qu'elles permettent ou non de remplir. Selon Brigandt, cette approche s'oppose aux théories récentes des espèces naturelles qui visent principalement à décrire la teneur ontologique des espèces naturelles. Son analyse approfondie du rapport entre objectifs et catégories lui permet ensuite de défendre l'abandon de l'étiquette « espèce naturelle ». En effet, son analyse, en accord avec le cohérentisme historique, met de l'avant la spécificité du contexte sociohistorique au sein duquel une catégorie est utilisée. Les catégories scientifiques apparaissent dès lors comme des entités mouvantes, comme le fruit de conjonctures particulières et non comme des entités naturelles (à moins, bien sûr, de soutenir une acception déflationniste du terme « naturelle » qui rendrait ce qualificatif trivial). Explorons ces deux points tour à tour.

Bien que Brigandt rejette l'expression « espèce naturelle, » il insiste sur le besoin de traiter des espèces ou des catégories scientifiques (*kinds*) d'un point de vue épistémologique. Pour ce faire, il suggère un cadre méthodologique formé de trois composantes complémentaires.

La première composante stipule que les catégories scientifiques doivent être analysées par l'entremise du rapport qu'elles entretiennent avec des objectifs humains. Cela permet d'expliquer comment une catégorie acquiert sa pertinence, son utilité étant relative aux objectifs de recherche (par exemple, les isotopes peuvent former des catégories inutiles pour la chimie, mais utiles pour certaines branches de la physique). De surcroît, appréhender les catégories sur cette base permet de discriminer plusieurs versions théoriques d'une même catégorie afin de conserver la plus efficace relativement à un objectif donné (par exemple, certaines articulations du concept d'homologie sont trop inclusives pour servir de matière aux inférences

phylogénétiques, de sorte que le sens que d'influents cladistes [comme Patterson, Nelson ou Williams] ont donné au terme [synonyme de synapomorphie] soit nécessaire pour répondre aux objectifs de la discipline). Bien que d'autres auteurs et autrices (Ereshefsky et Reydon 2015; Franklin-Hall 2015; Slater 2015) aient auparavant noté l'importance de penser les espèces naturelles en lien avec les objectifs auxquels elles répondent, l'originalité de l'approche de Brigandt est d'en faire l'axe d'analyse principal relativement aux catégories scientifiques.

Deuxièmement, l'analyse des pratiques classificatoires doit être précise (*fine grained*), ce qui contraste avec le niveau de généralité typiquement exploré par les théories des espèces naturelles. D'après Brigandt, le niveau de généralité requis par le fait même de viser une théorie unitaire des espèces naturelles force les théoriciens et théoriciennes à traiter de domaines ou de champs disciplinaires entiers comme des unités d'analyse pertinentes. Brigandt argumente que différentes catégories non concordantes peuvent pourtant servir les objectifs différents de domaines de recherche plus précis. L'analyse ciblant les discours de ces domaines permet ainsi de considérer des objectifs scientifiques plus concrets. Plutôt que de référer à des objectifs comme « l'explication » ou « l'induction », Brigandt suggère d'explorer l'arrimage de catégories scientifiques et d'objectifs comme « reconstruire un arbre phylogénétique » ou « évaluer l'impact de la racialisation sur diverses variables socioéconomiques ». De telles analyses porteraient plus directement sur le travail réellement effectué par les praticiens et praticiennes.

La troisième composante du cadre méthodologique suggéré par Brigandt consiste à insister sur les objectifs traditionnellement considérés comme étant non épistémiques et leur influence sur le choix et l'utilisation de catégories en sciences. En ce sens, Brigandt importe des considérations du luxuriant complexe discursif de philosophie, de sociologie et d'anthropologie des sciences portant sur l'influence des valeurs en sciences (par exemple, Elliott 2011, 2017; Intemann 2015; Longino 1990, 2016; Potochnik 2015; Winsberg 2012) afin de l'appliquer spécifiquement à l'analyse des pratiques classificatoires. Cela permet d'inclure des objectifs sociaux comme « mitiger l'impact des changements climatiques » ou « lutter contre le racisme » dans l'analyse du rôle des catégories étudiées au sein des systèmes épistémiques qui les manipulent. Une catégorie qui, d'un point de vue strictement biologique, pourrait sembler obsolète peut acquérir une pertinence renouvelée si elle permet de lutter contre la stigmatisation

sociale, par exemple. Se limiter aux objectifs traditionnellement considérés comme étant épistémiques serait donc trop contraignant.

Ces trois composantes jumelées permettent de saisir l'essence de l'approche suggérée par Brigandt. Selon lui, les pratiques classificatoires gagnent à être étudiées épistémologiquement à l'aune des interactions entre catégories et objectifs précis (épistémiques ou non) :

Generally, a combination of epistemic and nonepistemic kinds may motivate the use of a kind in scientific practice, which needs to be taken into account to assess the most appropriate construal of the kind. The cases that I discussed across sections 4 and 5 have illustrated the relevance of all three components, some of which other perspectives on scientific kinds may miss. The case discussion in particular addressed how nonepistemic aims, especially social-political aims, matter when considering how to construe the boundaries and nature of a kind. (Brigandt 2020, 17)

Cela étant dit, cette méthode pourrait être compatible avec la conservation de la notion d'espèce naturelle. En effet, elle pourrait servir à étudier le rôle épistémique que jouent les espèces naturelles identifiées ailleurs. Or, Brigandt rejette cette option. Il exhorte en effet les philosophes à traiter des catégories scientifiques en évitant d'utiliser l'expression « espèce naturelle », une injonction doublement motivée.

Premièrement, le qualificatif « naturelle » suggère que les catégories ou espèces sont des composantes d'un ordre naturel qui existerait indépendamment de notre cognition, en cohérence avec la tendance réaliste à l'égard des espèces naturelles présente au sein de la tradition. Or, les catégories scientifiques sont codéterminées par le monde empirique et les objectifs humains. De ce fait, l'utilisation du prédicat « naturelles » pour décrire ces catégories biaise le rapport que la philosophie de la classification entretient envers son objet d'étude. Ce prédicat dissimule une partie significative des facteurs qui déterminent les pratiques classificatoires et qui doivent donc être appréhendés pour bien saisir les dynamiques en jeu.

Deuxièmement, l'expression « espèce naturelle » suggère que les espèces candidates détiennent quelque chose comme une base physico-chimique immuable, une suggestion erronée puisque lorsque des bases physico-chimiques permettent effectivement de circonscrire certaines catégories, celles-ci ne sont que partiellement et approximativement déterminées par de telles bases. Brigandt donne l'exemple de la notion de « race » qui, chez les êtres humains, est

certainement liée à des traits biologiques, mais dont la portée et l'utilisation dépassent largement ces caractéristiques. Cet exemple est particulièrement porteur puisqu'il témoigne du fait que les faux positifs relatifs à l'utilisation du prédicat « naturelles » peuvent avoir des conséquences néfastes. Ces conséquences sont directement liées à la naturalité ainsi suggérée : « scholars should avoid using words that have connotations that obscure that many kinds reflect social processes and thus are ultimately subject to human responsibility. This clearly holds for the examples of race, gender, and psychiatric kinds discussed, and even for species as kinds. » (Brigandt 2020, 20).

En résumé, l'approche de Brigandt remet en question l'un des piliers de l'arche des espèces naturelles. Selon lui, il y a très certainement des catégories scientifiques qui méritent notre attention en raison du rôle épistémique qu'elles jouent en sciences. Il est donc important de se doter d'un cadre méthodologique pour les étudier, et même d'un cadre unitaire comme celui suggéré par l'auteur dans son article de 2020. Cependant, Brigandt rejette la seconde croyance de base de la tradition des espèces naturelles selon laquelle des catégories scientifiques détiendraient un statut métaphysique privilégié (traduit par l'utilisation de l'adjectif « naturel ») que les théories philosophiques de la classification se doivent d'appréhender. Au contraire, suggère Brigandt, les approches épistémologiques des classifications scientifiques devraient s'intéresser au rôle épistémique que les catégories scientifiques jouent, un rôle qui ne peut être compris que par une étude des objectifs (épistémiques ou sociaux) auxquels les catégories répondent ou non, et ainsi évaluer leur validité. Cette validité est en somme un phénomène culturel plutôt que naturel.

4.2.1.3 Les approches éliminativistes de Brigandt et de Ludwig, en bref, ainsi qu'une première esquisse de la mienne

Les deux arguments éliminativistes tout juste présentés s'attaquent à des éléments distincts de l'arche des espèces naturelles. Ludwig remet en question la possibilité de faire reposer l'arc (la croyance selon laquelle une ou des théories des espèces naturelles seraient nécessaires à la compréhension des pratiques classificatoires en sciences) sur les deux piliers (croyances selon lesquelles les espèces naturelles ont des statuts particuliers sur les plans métaphysique et épistémique). Selon lui, la diversité des catégories scientifiques est irréductible

et invite à abandonner le concept général d'espèce naturelle au profit d'une approche multidimensionnelle des classifications non arbitraires. Brigandt, quant à lui, considère que certaines catégories scientifiques détiennent un rôle épistémique particulier, ce qui justifie effectivement le développement d'une cadre d'analyse particulier pour les appréhender. Cependant, le tout serait indépendant de la croyance selon laquelle ces catégories auraient un statut métaphysique particulier, une « naturalité » quelconque. Les deux auteurs ont donc en commun de vouloir rejeter l'utilisation de l'expression « espèce naturelle » et son rôle central au sein des pratiques philosophiques cherchant à rendre compte des pratiques classificatoires en sciences.

Il est aussi très important de souligner que Ludwig et Brigandt, en plus d'étayer une critique de l'arche des espèces naturelles (à leur insu), offrent des outils analytiques de rechange. Pour Ludwig, il s'agit simplement de profiter de la grande diversité des écrits au sein de la tradition des espèces naturelles pour analyser de façon détaillée les pratiques classificatoires. Pour Brigandt, il faut appréhender les catégories scientifiques en s'intéressant aux objectifs humains auxquels elles peuvent répondre (ou non).

À ce moment-ci, il est important de noter que d'autres auteurs, au sein même de la tradition des espèces naturelles, ont mis de l'avant des méthodes d'analyse qui sont tout à fait compatibles avec un rejet de la notion en question. Je pense notamment aux travaux d'Ereshefsky et Reydon (2015) ainsi qu'à ceux de Chang (2012, 2016, 2017). Ces auteurs appréhendent la classification par l'entremise de programmes classificatoires et des dynamiques socioculturelles qui font respectivement émerger et disparaître certaines catégories. Dans les deux cas, l'étiquette « espèce naturelle » est accessoire à l'analyse des pratiques classificatoires mises de l'avant. Faute d'espace et de temps, je dois faire ici l'économie de la démonstration exhaustive de ce constat. L'important est simplement de souligner, encore une fois, que le rejet de la notion d'espèce naturelle permet tout de même de conserver les outils philosophiques de qualité ayant été développés au sein de la tradition qui fait l'objet de ma critique.

Avant de compléter la revue du corpus relatif aux espèces naturelles en explorant un argument contre l'éliminativisme (Conix et Chi 2021), il convient de fournir une première esquisse

de ma posture éliminativiste afin de souligner en quoi elle se distingue des deux approches tout juste présentées (Brigandt 2020; Ludwig 2018). Alors que ces deux approches appréhendent de façon critique certains éléments de l'arche des espèces naturelles, je remets en question l'ensemble de cette structure épistémologique. Plus spécifiquement, je considère que l'épistémologie des pratiques classificatoires gagne à éviter la notion d'espèce naturelle pour trois raisons : premièrement, la notion en question est théoriquement inutile (fait reconnu notamment par Conix et Chi); deuxièmement, les théories des espèces naturelles sont plutôt des mécanismes de légitimation des pratiques classificatoires que des théories utiles pour l'analyse épistémologique de ces pratiques (Hacking 1992, 1994, 2009); troisièmement, la tradition des espèces naturelles suggère un point focal inadéquat, soit les catégories scientifiques, pour les analyses des pratiques classificatoires. Ce dernier point témoigne du fait que mon argument va plus loin que ceux de Ludwig et Brigandt puisque je suggère qu'il est non seulement problématique de centrer les analyses épistémologiques des pratiques classificatoires sur les espèces naturelles, mais aussi sur les catégories en général. Par ailleurs, le fait d'abandonner la notion d'espèce naturelle est peu coûteux puisqu'un cadre d'analyse de rechange efficace est offert par le cohérentisme historique (en plus des éléments mis de l'avant par Brigandt et Ludwig). L'ensemble de cette posture argumentative est détaillé dans la deuxième partie de ce chapitre. Ma réponse à l'argument de Conix et Chi, en 4.2.2, fait office de mise en bouche pour le plat de résistance qu'est mon argument éliminativiste étayé en 4.3.

4.2.2 Un argument contre l'éliminativisme et ma réponse à cet argument

Dans cette section, j'explore un argument explicitement formulé pour répondre à la posture éliminativiste du Ludwig. Je démontre que l'argument de Conix et Chi (2021), bien qu'il soulève des éléments importants à considérer pour la mise en place d'une approche philosophique des pratiques classificatoires, est insuffisant pour soutenir la nécessité de la notion d'espèce naturelle.

Par une riche analyse du rôle épistémique des espèces naturelles publiée en 2021, Conix et Chi argumentent que le pluralisme à l'égard des espèces naturelles, bien qu'il mette le doigt sur la faible valeur théorique de la notion (Hacking 2007; Ludwig 2018), est insuffisant pour

justifier une posture éliminativiste. Leur texte souligne en effet que l'expression « espèce naturelle » peut jouer différents rôles pertinents au sein du milieu académique, même s'il est impossible de formuler une théorie unitaire concernant les espèces naturelles et *même si cette notion est théoriquement inutile* (Conix et Chi 2021).

Leur analyse débute par un retour sur l'argument pluraliste et éliminativiste à l'égard des espèces naturelles, comme formulé par Hacking et plus récemment par Ludwig. Selon Conix et Chi, ces auteurs affirment que l'expression « espèce naturelle » est non seulement inutile, mais qu'elle est de surcroît nuisible. Pour rappel, Ludwig soutient en effet que la tentative de formuler une théorie unique concernant les espèces naturelles entraîne une perte de précision analytique.

En réponse, Conix et Chi défendent la double utilité de la notion d'espèce naturelle. D'abord, cette notion peut être considérée comme utile en tant que concept d'investigation (*investigative concept*). Ensuite, l'expression « espèce naturelle » sert de liant pour la communauté pluridisciplinaire de chercheurs et de chercheuses s'intéressant aux aspects philosophiques des pratiques classificatoires. Une posture éliminativiste, pour se faire convaincante, doit démontrer que l'utilisation du concept est plus nuisible que bénéfique; il faut donc tenir compte de tous les rôles potentiels du concept et non seulement du rôle théorique qu'un concept général d'espèce naturelle pourrait ou non remplir. Dans ce qui suit, j'analyse tour à tour les deux rôles du concept d'espèce naturelle et je remets en question la validité d'une justification de l'utilisation de ce concept basée sur ces rôles.

Les espèces naturelles remplissent un rôle au sein de la recherche que Conix et Chi décrivent par la notion de concept d'investigation. Un concept d'investigation est une notion qui aide à circonscrire un objet d'étude, un phénomène, sans pour autant être mobilisée au sein des explications ou des théories qui traitent du phénomène en question (comme le concept de vie en biologie ou celui d'être humain en sciences sociales). Selon Conix et Chi, le concept d'espèce naturelle permet ainsi de circonscrire une famille de phénomènes d'intérêt pour l'épistémologie ou la philosophie des sciences :

More precisely, we argue that philosophers of science use it to refer to a phenomenon of interest in scientific classificatory practices. MacLeod et Reydon (2013a, p. 90) point to this phenomenon when they write that 'at a very basic level scientists, but also

philosophers, share a need to distinguish between natural—with connotations such as objective, existing “out there” in the world, real, stable, unique, etc.—and artificial ways of grouping things together’. That is, scientists in a wide variety of disciplines seem to consider some kinds reflections of the world, and other kinds merely pragmatic tools. (Conix et Chi 2021, 9003)

Conix et Chi appellent « phénomène des espèces naturelles » (*natural kinds phenomena*) le fait que les philosophes et les scientifiques cherchent à distinguer, parmi les catégories scientifiques, celles qui sont naturelles de celles qui sont artificielles. La croyance selon laquelle une telle distinction existe est si bien enracinée au sein des pratiques philosophiques et scientifiques qu’il s’agit, si on accepte l’argument de Conix et Chi, d’un phénomène central aux pratiques classificatoires duquel il faudrait que l’épistémologie rende compte.

Conix et Chi embrassent la dimension pluraliste de l’argument éliminativiste de Ludwig. Les pratiques classificatoires sont multiples et les catégories candidates au statut d’espèce naturelle le sont aussi; rien ne permet de croire qu’une théorie unificatrice des espèces naturelles permettrait de rendre compte de toute cette diversité. Le concept étant ainsi fracturé, son utilité théorique est indubitablement limitée. Cependant, il joue un rôle d’investigation qui détient deux aspects complémentaires.

Premièrement, le concept d’espèce naturelle incite à tracer des parallèles entre différentes théories des espèces naturelles de sorte à expliquer la source de ces ressemblances (qui existent malgré la différence). Dans cette optique, le concept d’espèce naturelle souligne seulement le fait que deux théories distinctes traitent du même phénomène :

It is clear that the role of NK concept in such research is not theoretical. Clarifying the overlap between different accounts of natural kinds does not require us to define natural kinds or restrict the usage of that concept to a particular subset of scientific kinds. [...] The general concept of natural kinds then serves as the framework within which different kinds of natural kinds are conceptualised as part of the same cluster of phenomena. (Conix et Chi 2021, 9007)

Deuxièmement, la notion d’espèce naturelle, en unifiant le champ d’investigation, permet aussi de faciliter les constats généraux à l’égard des théories des espèces naturelles, les traits communs à ces théories pouvant engendrer des critiques sur l’ensemble du complexe discursif et ainsi mener à son amélioration. Par exemple, il pourrait être suggéré, comme Brigandt (2020) le fait, que les théories des espèces naturelles ont tendance à être formulées à un haut niveau

d'abstraction et que cela écarte des informations importantes par rapport aux phénomènes que sont les pratiques classificatoires.

En bref, le rôle d'investigation de la notion d'espèce naturelle découle de l'identification du « phénomène des espèces naturelles, » soit le fait que les praticiens et praticiennes considèrent que certaines catégories scientifiques sont plus naturelles (réelles) que d'autres. Les théories des espèces naturelles seraient ainsi unifiées par leurs tentatives d'explication du même phénomène. Cette unification rendrait possibles et constructives des analyses comparatives portant sur ces théories, leurs traits partagés, leurs forces et leurs faiblesses, ou autrement dit, leur inclusion au sein d'un seul et même complexe discursif.

L'argument basé sur le rôle d'investigation du concept d'espèce naturelle, en tant que réponse à l'éliminativisme, contourne l'enjeu central. L'argument de Conix et Chi est basé sur le phénomène des espèces naturelles et les avantages liés au fait de structurer l'épistémologie autour de la croyance sous-jacente selon laquelle il est important de distinguer espèces naturelles et artificielles : « Understanding classification in science, as philosophers aim to do, requires making sense of this 'natural kind phenomenon'. » (Conix et Chi 2021, 9003). Or, cette logique met sur un piédestal l'une des croyances des académiciens et des académiciennes. La question se pose cependant : cette croyance des philosophes et des scientifiques selon laquelle certaines catégories seraient plus naturelles que d'autres est-elle la seule qui figure au sein des systèmes épistémiques qui sous-tendent des pratiques classificatoires? Si ce n'est pas le cas (ce n'est pas le cas), pourquoi prioriser cette croyance afin de guider nos analyses épistémologiques des pratiques classificatoires? Pourquoi analyser cette croyance plutôt que, par exemple, celles selon lesquelles le modèle arborescent, les hiérarchies classificatoires, le monisme, le lien similarité-phylogénie, la définition de l'espèce biologique, le fait même de catégoriser des individus, etc. sont des principes épistémiques adéquats?

C'est pourtant bien ce point qui est au cœur des postures éliminativistes. Ludwig considère que le phénomène des espèces naturelles est certes intéressant en soi (l'ethnobiologie fait la distinction entre les catégories que les êtres humains considèrent comme détenant un statut ontologique fort de celles qui sont considérées plus pragmatiquement), mais

il ne s'agit là que d'un axe d'analyse parmi d'autres tout aussi pertinents. Cet axe devient nuisible s'il est priorisé *de facto*. Cela encourage en effet le développement de théories générales plutôt que de théories plus précises. L'argument de Brigandt fait aussi écho à ce constat en invitant à réaliser des analyses plus fines. Je dois insister sur ce point : il y a de bonnes raisons de croire que l'analyse philosophique des pratiques classificatoires devrait éviter de se limiter à l'analyse du phénomène des espèces naturelles ou même, selon moi, des catégories scientifiques. Le phénomène et le concept correspondant seraient dès lors relégués à la périphérie des analyses des pratiques classificatoires; Ludwig et Brigandt (et la deuxième partie de ce chapitre) offrent des arguments convaincants pour soutenir cette thèse. Plutôt que d'y répondre de front, Conix et Chi acceptent de fait la centralité du phénomène des espèces naturelles.

En résumé, le premier argument de Conix et Chi stipule qu'il serait néfaste d'écarter le concept d'espèce naturelle puisque celui-ci joue un rôle d'investigation important en circonscrivant le phénomène auquel devrait s'intéresser les analyses épistémologiques de la classification. Ma réponse à cet argument est que cela crée un biais injustifié envers l'une des croyances constitutives des systèmes épistémiques liés aux pratiques classificatoires. Ce biais tend à occulter la complexité du phénomène que sont les pratiques classificatoires. Autrement dit, même si les scientifiques considèrent que certaines de leurs catégories sont plus naturelles que d'autres, les philosophes des sciences devraient éviter d'appuyer leurs analyses sur une telle croyance. Cette croyance devrait être présentée comme un objet d'étude parmi d'autres plutôt qu'utilisée comme principe structurant des cadres d'analyse philosophiques.

Le deuxième argument de Conix et Chi se rejette plus simplement. Selon cet auteur et cette autrice, l'expression « espèce naturelle » joue un double rôle social en plus de son double rôle en tant que concept d'investigation. Premièrement, cette expression lie entre eux et elles un ensemble de chercheurs et de chercheuses en philosophie. Malgré la diversité des travaux qui réfèrent à cette notion, Conix et Chi soulignent qu'ils témoignent d'une unité au sein de la communauté qui s'exprime par le désir partagé d'appréhender les pratiques classificatoires par

la même porte d'entrée²⁵ : « Diverse as these contributions may be, their association to the phrase makes them into a connected body of work that any philosopher of science working on scientific classification will start from and engage with. » (Conix et Chi 2021, 9008).

Deuxièmement, l'expression « espèce naturelle » lie cette communauté de recherche à d'autres communautés à l'extérieur de la philosophie. Les biologistes ou chimistes s'intéressant à l'aspect philosophique des pratiques classificatoires, par exemple, utiliseraient cette expression pour guider leur exploration des thématiques concernées.

Ces deux aspects du rôle social de l'expression « espèce naturelle » sont soutenus par des données bibliométriques offertes par Conix et Chi, qui en concluent que l'expression en question est indispensable : « the phrase facilitates research by tying together a research community and research tradition. As eliminating natural kinds from philosophy would get rid of this phrase, *eliminativism comes at the cost of losing the social role of the phrase too.* » (Conix et Chi 2021, 9008, italiques ajoutés). Cette conclusion est beaucoup trop forte. En effet, tout porte à croire qu'une autre expression relative à la classification, comme l'expression « pratiques classificatoires » pourrait en venir à jouer le même rôle social dans le futur.

De plus, ni les postures éliminativistes explorées plus tôt ni la mienne ne demandent aux praticiens et aux praticiennes de la philosophie des sciences d'effectuer une transition immédiate et drastique qui larguerait dans l'abîme de l'oubli l'ensemble des travaux ayant été réalisés au sein de la tradition des espèces naturelles. Au contraire, il est explicitement recommandé par Ludwig (comme souligné précédemment) de mobiliser les outils qui y ont été développés. Le tissu bibliométrique de la communauté philosophique, efficacement documenté par Conix et Chi, se maintiendrait probablement par référence aux textes de la tradition, en plus d'être enrichi par l'importation de réflexions connexes, comme le fait Brigandt en référant aux travaux portant sur l'influence des valeurs en sciences (l'utilisation du cohérentisme historique, développé en philosophie de la mesure, relève aussi de ce type d'importation). Dans un tel contexte, la transition graduelle écartant l'utilisation du terme « espèce naturelle » ne viendrait aucunement

²⁵ Bien que cet aspect social du phénomène des espèces naturelles chevauche selon moi son rôle d'investigation, Conix et Chi insistent pour distinguer les deux.

nuire à la cohésion de la communauté des philosophes qui se construirait autour d'une nouvelle étiquette (comme « pratiques classificatoires ») évoquant l'objet d'étude partagé, une étiquette que les scientifiques à l'extérieur de la communauté sauraient probablement utiliser pour guider leurs interrogations philosophiques, le moment venu (j'invite en effet le lecteur et la lectrice à avoir un minimum de confiance envers la capacité des chercheurs et des chercheuses à explorer un corpus de textes qui les intéresse même si celui-ci est caractérisé par une transition conceptuelle).

Une autre façon d'illustrer la force excessive de la conclusion de Conix et Chi est de l'appliquer à un autre objet d'étude. Ayant reconnu l'inutilité théorique d'une notion (comme Conix et Chi le font pour celle d'espèce naturelle), est-il sage de continuer de structurer un champ disciplinaire autour de celle-ci? Serait-il pertinent de penser la classification biologique autour de la notion de création divine ou la description des corps physiques sur la base du géocentrisme? Poser la question, c'est y répondre. Ce point peut être reçu comme une réponse à l'argument de Conix et Chi qui repose sur le rôle social de la notion d'espèce naturelle, mais aussi à l'argument basé sur son rôle d'investigation.

En résumé, l'argument de Conix et Chi contre l'éliminativisme stipule que les rôles sociaux et d'investigation de la notion d'espèce naturelle contrebalancent les aspects nuisibles de la notion et son inutilité théorique (Brigandt 2020; Ludwig 2018). Il serait ainsi justifié de la conserver et d'en faire l'étoile Polaire (ou la Croix du Sud) des analyses philosophiques des pratiques classificatoires en sciences. En réponse, j'ai suggéré que le rôle social peut être rempli par une expression de rechange comme « pratiques classificatoires » qui maintiendrait le tissu social de la communauté des chercheurs et des chercheuses s'intéressant aux aspects philosophiques des pratiques classificatoires.

Plus importante est ma réponse à l'utilité investigatrice de la notion d'espèce naturelle. Cette réponse est au cœur de l'approche éliminativiste que je cherche à mettre de l'avant et à défendre concrètement dans la prochaine partie du chapitre. Il est néfaste de structurer la recherche portant sur les pratiques classificatoires autour de la croyance selon laquelle il y a une différence entre espèces naturelles et artificielles, une croyance qui résume l'arche des espèces

naturelles. Cette pratique est problématique puisque cela suggère qu'un regard centré sur un élément des pratiques classificatoires, soit la conception que les praticiens et praticiennes se font de leurs catégories, permet de rendre compte de l'ensemble des pratiques étudiées. Or, ces catégories n'acquièrent leur sens, aux yeux des praticiens et praticiennes, que par l'entremise d'un système épistémique complexe souvent implicite à leur discours. Afin de saisir les dynamiques scientifiques relatives à la classification, il faut rendre compte de l'évolution de tels systèmes et non simplement de l'une des croyances qui les structurent.

4.3 De l'utilité épistémologique de la notion d'espèce naturelle

Dans cette partie, j'articule mon argument pour le rejet de la notion d'espèce naturelle comme élément structurant d'une épistémologie de la classification en mobilisant comme exemple la notion d'homologie. En 4.3.1, j'effectue un bref rappel du rôle qu'a occupé et qu'occupe actuellement la notion d'homologie au sein de la phylogénétique, en plus de présenter brièvement des descriptions faites de l'homologie en termes d'espèces naturelles. En 4.3.2, j'explore et réfute un argument de Brigandt (2009) qui défend explicitement l'utilité épistémologique d'une description de l'homologie en tant qu'espèce naturelle au sens que Boyd (1999a) donne à cette expression (groupe homéostatique de propriétés). En 4.3.3, je présente deux arguments auxiliaires qui renforcent mon rejet de l'utilité épistémologique de la notion d'espèce naturelle.

4.3.1 L'homologie phylogénétique et les groupes de propriétés homéostatiques

La notion d'homologie est centrale à l'ensemble de la biologie comparative. De ce fait, elle a fait couler beaucoup d'encre au travers du temps (et même au travers de cette thèse). Elle a aussi été fractionnée et précisée de sorte à identifier des relations toujours plus strictes entre traits biologiques. Cette multiplication des acceptations associées à la notion d'homologie (Minelli et Fusco 2013) peut être illustrée par la biologie développementale, qui a insisté sur la teneur ontogénique que peut détenir le concept avant de réconcilier cette définition avec la perspective phylogénétique sur la question (Assis et Brigandt 2009; Brigandt 2003, 2007; Brigandt et Griffiths 2007; Griffiths 2007; Müller 2003; Pavlinov 2012; Wagner 1989, 1994, 1996). La pluralité des

conceptions de l'homologie transparaît aussi dans l'opposition entre l'homologie taxique et transformationnelle (Assis 2013; Brower 2015; Brower et de Pinna 2012; Brigandt 2003; Carine et Scotland 1999; Love 2007; Nelson 1994; Nixon et Carpenter 2012; Patterson 1982; Rieppel 2015; Van Valen 1982; Williams et Ebach 2012) ou encore dans les débats entourant le statut métaphysique des homologies, pensées comme étant des individus, des espèces naturelles ou encore les deux (Brigandt et Griffiths 2007; Griffiths 1999; Kluge 1989, 1997, 2003; Pearson 2019; Rieppel 1994, 2015; Wilson 1996). Dans le cadre de ce chapitre, j'évite de prendre position par rapport à ces débats qui traitent plus spécifiquement d'aspects métaphysiques et théoriques de la biologie que d'épistémologie. Je cherche uniquement à tester l'hypothèse selon laquelle une description de l'homologie en termes d'espèces naturelles peut nous aider à comprendre les pratiques phylogénétiques.

Nonobstant les débats mentionnés, un consensus a émergé relativement à la définition de l'homologie à un plus haut niveau d'abstraction. Celle-ci réfère à la relation entre deux traits qui partagent un passé évolutif, c'est-à-dire qu'il s'agit de deux versions du « même » trait ayant existé chez un ancêtre commun (Inkpen et Doolittle 2016; Pavlinov 2012). Ainsi, à ce niveau d'abstraction, l'essence de l'homologie a été conservée depuis son origine (Owen 1843) : elle réfère à un lien de similarité considéré comme étant théoriquement pertinent (ce qui implique, dans le contexte qui m'intéresse, une pertinence évolutive) et qui rend légitimes les inférences de la biologie comparative.

Il est courant de décrire les homologies ou les regroupements d'homologues²⁶ comme des espèces naturelles, et plus précisément, comme des espèces naturelles GHP (Assis et Brigandt 2009; Griffiths 2007; Keller, Boyd et Wheeler 2003; Pearson 2019; Rieppel 2005a, 2015; Wagner 1996). L'idée générale est la suivante : malgré le fait qu'ils soient présents sous différentes formes, deux traits peuvent être considérés comme relevant du « même » caractère parce qu'ils

²⁶ L'expression « les homologies sont des espèces naturelles » et « les regroupements d'homologues sont des espèces naturelles » ne sont pas équivalentes puisque l'homologie est la relation entre les homologues. Pourtant, le corpus étudié dans ce chapitre semble les considérer comme synonymes puisque le statut d'espèce naturelle des homologies est débattu sur la base de la similarité entre les homologues qui seraient les occurrences de l'espèce naturelle concernée. Je n'endorsse *pas* cette équivalence, mais j'essaie de présenter les arguments avec fidélité.

sont composés de propriétés qui ont tendance à être coréalisées *et* parce que cette cooccurrence est expliquée par un ou des mécanismes homéostatiques. Par exemple, tous les membres avant des tétrapodes sont homologues parce qu'ils partagent certaines propriétés (mêmes os présents et organisés de la même façon, même position par rapport au reste du corps, etc.) et que cette similarité est expliquée par des mécanismes sous-jacents (hérédité, brassage génétique au sein des lignées, mécanismes développementaux, etc.). Ainsi, une catégorie formée de toutes les versions d'un caractère, s'il s'agit bien dans tous les cas du *même* caractère, serait une homologie *et* une espèce naturelle au sens que Boyd a donné à cette notion (Assis et Brigandt 2009; Brigandt 2009; Brigandt et Griffiths 2007; Pearson 2019; Rieppel 2015).

Je n'ai aucun doute relativement à la possibilité d'établir l'adéquation entre cette théorie des espèces naturelles et la notion d'homologie (les auteurs cités traitent principalement de l'homologie sur les plans morphologique et développemental, mais un tel arrimage avec le niveau moléculaire est sûrement possible). Plutôt que de remettre en question la validité de la théorie en question, mon objectif ici est d'interroger l'utilité du constat selon lequel une homologie est une espèce naturelle GHP dans le cadre d'une épistémologie de la phylogénétique. Pour ce faire, il convient de se pencher sur un texte de Brigandt de 2009 dans lequel cet auteur défend explicitement l'utilité de parler de l'homologie par l'entremise d'une théorie des espèces naturelles, une suggestion régulièrement prise pour acquise au sein de la tradition des espèces naturelles (il va sans dire que j'analyserai le texte en question et non la pensée de Brigandt, qui s'est distancié de la notion d'espèce naturelle depuis 2009).

4.3.2 Brigandt et l'utilité épistémologique de la notion d'espèce naturelle

Dans son texte de 2009, Brigandt suggère que la correspondance entre une catégorie et la définition d'une espèce naturelle GHP est très peu informative puisqu'une grande variété de catégories scientifiques, très différentes les unes des autres, correspondent à cette définition. Cependant, le processus qui mène à l'établissement d'une telle correspondance est quant à lui épistémologiquement révélateur :

The flipside of this is that the very claim *that* a collection of objects forms an HPC natural kind is relatively weak and as such not informative. The real question is *how and for what reason* these objects form an HPC kind; in particular, what the particular

properties that are fairly reliably correlated are, and what the specific homeostatic mechanisms forming the causal basis for this correlation and the unity of the kind are. Knowing about these concrete features characterizing a particular natural kind permits scientific inferences and explanations [...]. The very *metaphysical* fact that something is a natural kind implies that there are properties that (if known) can be used for the purposes of scientific induction and explanation, yet such *epistemic* aims are actually met only to the extent to which there is a natural kind *concept* that embodies knowledge about the scientifically relevant features of the kind. (Brigandt 2009, 93, italiques présents dans le texte original)

Cette citation nous permet d'identifier deux points saillants de l'argument de Brigandt. Selon lui, l'identification d'une espèce naturelle GHP permet de mettre le doigt sur un fait métaphysique qui sous-tend les propriétés corrélées et qui témoigne de l'unité de la catégorie. Ce serait ce fait métaphysique qui implique la présence de propriétés corrélées qui peuvent être utilisées dans le cadre d'inférences et d'explications scientifiques. De plus, nous dit Brigandt, les objectifs épistémiques (réalisation d'inférences et d'explications) ne peuvent être réalisés que si nous avons des concepts, comme celui d'homologie, qui correspondent à des espèces naturelles. Parler d'espèces naturelles, en épistémologie, serait donc utile pour deux raisons : d'abord parce que la métaphysique ainsi décrite explique les succès épistémiques des pratiques scientifiques, et ensuite parce que de tels succès ne peuvent être atteints que si nous avons des concepts correspondant à des espèces naturelles qui incarnent nos connaissances.

Dans le but de répondre à ces deux aspects de l'argument de Brigandt, considérons les grandes lignes d'un processus d'inférence phylogénétique typique effectué dans une optique cladiste et mobilisant la méthode de parcimonie afin d'identifier le rôle que l'homologie y occupe (Assis 2015; Farris 1983; Kearney et Rieppel 2006; Mooi et Gill 2010). Pour rappel, l'homologie est dans ce contexte-ci synonyme de synapomorphie. Je tiens à souligner un autre point : aucun des auteurs et des autrices cité-e-s n'utilise la notion d'espèce naturelle pour décrire le processus inférentiel duquel il est question.

La première étape, selon les auteurs et autrices cité-e-s, consiste à formuler des hypothèses d'homologie. Celles-ci requièrent l'observation de traits morphologiques (j'explore le traitement de données moléculaires au prochain chapitre), de caractères existant sous différents états au sein d'unités taxonomiques préalablement formées (sur la base de la similarité des membres). La formulation d'hypothèses d'homologie est alors réalisée sur la base des critères de

similarité traditionnels (structure du trait, emplacement, présence d'états de caractère intermédiaires, etc.) caractérisant la « bonne » similarité reliant deux états de caractère, celle qui a de fortes chances de relever de l'ascendance commune des organismes qui portent les potentiels homologues (Assis 2015; Mooi et Gill 2010). Ces premières hypothèses d'homologie peuvent être triées sur la base d'études développementales, morphologiques, fonctionnelles, structurelles, etc. subséquentes (Hennig 1966; Kearney et Rieppel 2006; Rieppel 2005b, 2009) ou non (Grant et Kluge 2004; Kluge 2003; O'Leary, Gatesy et Novacek 2003; Patterson 1982). Dans les deux cas, c'est ultimement la congruence qui permet de confirmer ou d'infirmer une hypothèse d'homologie (Assis 2015; Kluge 1998; Patterson 1982; Rieppel 2015).

Pour tester la congruence des homologies, il faut reconstruire les relations phylogénétiques entre les entités concernées, ce qui représente la deuxième étape du processus. Les homologies qui soutiennent les mêmes topologies sont dites congruentes. Lorsque plusieurs phylogénies sont possibles et soutenues par les données, le principe de parcimonie invite à compter le nombre de changements d'états de caractères requis dans chaque scénario évolutif, puis à retenir celui ou ceux qui présentent le moins de changements d'états de caractères (Assis 2015; Farris 1983). Une fois le cladogramme représentant la phylogénie étant reconstruit, il peut être découpé de sorte à en extraire des catégories de rangs supérieurs à celui de l'espèce biologique sur la base de l'identification de clades ou de groupes monophylétiques (Panchen 1992). C'est ainsi qu'un système taxonomique et phylogénétique peut être érigé : par la mise en relations d'unités taxonomiques sur la base de liens de similarités postulés comme étant phylogénétiquement chargés, puis sur la représentation du système ainsi généré sous forme de catégories hiérarchiques.

Revenons maintenant aux deux arguments de Brigandt (2009). Premièrement, nous dit-il, le fait d'associer l'homologie aux espèces naturelles GHP permettrait d'identifier la cooccurrence de certaines propriétés et les mécanismes homéostatiques qui expliquent la corrélation entre ces propriétés. Son argument va cependant plus loin puisqu'il affirme que c'est l'existence d'une espèce naturelle qui explique la valeur épistémique des observations réalisées.

Or, les hypothèses d'homologie sont formulées *après* que la morphologie des organismes a été analysée pour tracer des liens de similarité. Si plusieurs traits analysés témoignent d'une structure osseuse et d'un placement (relativement au reste du corps) similaires au sein d'espèces différentes, il est alors possible de postuler l'homologie. La présence des os, leurs relations les uns aux autres et les relations topologiques de la structure par rapport au reste du corps sont alors des propriétés qui témoignent d'une certaine « homéostasie », c'est-à-dire qu'elles tendent à être réalisées ensemble. Ces observations mènent à l'inférence phylogénétique *que l'homologie est par la suite confirmée ou non par le test de congruence*. L'inférence de liens phylogénétiques, sur la base de liens de similarité précède donc logiquement l'identification de regroupement d'homologues.

C'est donc dire que le fait métaphysique sous-jacent, même s'il existe, est inconnu des praticiens et des praticiennes au moment de la réalisation des inférences. Il peut donc difficilement expliquer la réalisation d'inférences et d'explications. C'est plutôt parce que les praticiens et praticiennes mobilisent des procédures d'analyse de la similarité acceptées qu'ils et elles peuvent inférer des relations phylogénétiques. Il s'ensuit par ailleurs que l'homologie remplit son rôle inférentiel en tant que relation entre objets étudiés, relation de similarité théoriquement pertinente, *et non en tant que catégorie puisque la catégorie correspondante est identifiée après que le processus inférentiel mobilisant l'homologie a eu lieu*.

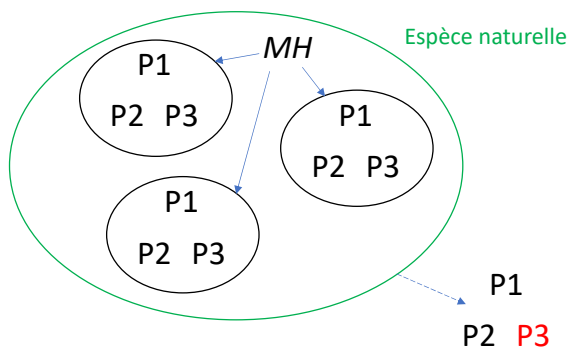
Ainsi, une description qualitative du processus inférentiel, comme celle qui précède, suffit à identifier les similarités considérées comme théoriquement pertinentes par les praticiens et praticiennes. Nul besoin, contrairement à ce que Brigandt (2009) suggère, d'identifier des espèces naturelles pour comprendre le comment et le pourquoi des inférences phylogénétiques mobilisant cette similarité. De surcroît, le postulat selon lequel l'homologie est une espèce naturelle ajoute une information qui ne reflète *pas* la structure des raisonnements phylogénétiques. En ce sens, je considère que ce postulat nuit à la précision des descriptions épistémologiques de la discipline en question.

Toujours par rapport au premier argument de Brigandt, un autre point qui dépasse le contexte de la phylogénétique doit être précisé. L'argument surestime largement l'importance

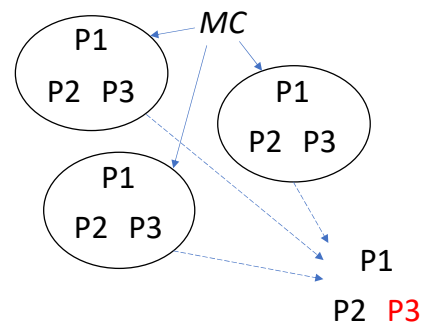
des catégories et de leur statut métaphysique dans le cadre d'inférences inductives : que les mécanismes observés et les propriétés coréalisées soient associés à une espèce naturelle ou non, cela ne change rien au pouvoir inférentiel et explicatif lié à l'observation empirique de stabilité. Considérons un exemple fictif. Les propriétés *P1*, *P2* et *P3* tendent à être coréalisées et associées au mécanisme *H*. Considérant ces informations, l'observation d'un phénomène présentant *P1* ou *P1* et *P2* ainsi que *H* suggère très fortement que *P3* est aussi présente (même si on ne l'a pas encore observée). La logique qui sous-tend la théorie GHP des espèces naturelles semble ici adéquate puisqu'un mécanisme et un groupe de propriétés sont liés de manière à permettre des inférences et à soutenir des explications. Or, le fait que les occurrences du groupe de propriétés *P1*, *P2* et *P3* forment une espèce naturelle ou non ne change rien à ce pouvoir inférentiel ou explicatif (voir Figure 4.1). C'est bien l'observation de stabilité qui rend l'inférence possible.

Figure 4.1 – Groupes de propriétés et inférences inductives

1.a. Inférence inductive avec espèce naturelle



1.b. Inférences inductives sans espèce naturelle



Cette figure illustre la possibilité de réaliser des inférences inductives à la suite de l'observation de groupes de propriétés corrélées. En 1.a, les groupes de propriétés semblent former une espèce naturelle sous-tendue par un mécanisme homéostatique. La stabilité de la corrélation permet d'inférer la présence de *P3* lorsque *P1* et *P2* sont observées (et *P3* non). En 1.b, les groupes de propriétés ne forment pas une espèce naturelle mais un mécanisme causal engendre une corrélation entre les propriétés de sorte que les inférences inductives sont disponibles. Comme détaillé dans le corps du texte, cette comparaison remet en question l'affirmation selon laquelle le fait métaphysique concernant une espèce naturelle est ce qui explique la possibilité de réaliser les inférences concernées.

Cela dit, décrire l'homologie en tant qu'espèce naturelle GHP pourrait tout de même être justifié si cela nous permet d'identifier des mécanismes homéostatiques qui expliquent la cooccurrence des propriétés des homologues qui auraient jusqu'alors été négligés. Dans le cas de la phylogénétique, les mécanismes qui assurent la stabilité des cooccurrences entre propriétés sont toutefois bien connus (hérédité, fluidité génétique au sein des populations, isolement

géographique des populations, spéciation, contraintes et mécanismes développementaux, etc.), de sorte qu'il est inutile de mobiliser une théorie des espèces naturelles pour les mettre à jour (ce point est particulièrement clair dans le cas de l'homologie moléculaire alors que les modèles mobilisés pour l'analyse de similarité sont explicitement de nature évolutive). De plus, si ces mécanismes étaient inconnus, il serait difficile de saisir comment une théorie des espèces naturelles pourrait nous aider à les mettre à jour. Ce travail est celui des praticiens et des praticiennes et non celui des épistémologues qui étudient leurs pratiques.

Le deuxième argument de Brigandt (2009) soutient que pour évaluer la capacité de la discipline (ici, la phylogénétique) à remplir ses objectifs, il faut s'intéresser aux espèces naturelles puisque les concepts qui leur sont associés sont cruciaux à la réalisation du processus inférentiel. Évidemment, si l'objectif des praticiens et des praticiennes est d'identifier des espèces naturelles, il faut définitivement se frotter à cette notion. Cependant, il faudrait les appréhender de la même manière qu'on s'intéresse à d'autres objets d'étude (croyances variées, modèles, définitions, concepts, etc.) ou objectifs de recherche (identifier des liens de similarité entre espèces ou des liens phylogénétiques entre unités taxonomiques opérationnelles, mettre à jour divers phénomènes évolutifs, mieux comprendre certains phénomènes physiologiques, etc.). Les espèces naturelles seraient autrement dit un élément interne aux pratiques classificatoires étudiées par l'épistémologie. De ce fait, il serait alors problématique d'utiliser comme outil d'analyse une notion qui constitue le phénomène analysé. Je reviens sur ce point en 4.3.3.

De toute façon, il est difficile de défendre que l'identification d'espèces naturelles soit un objectif central à la phylogénétique. Les pratiques phylogénétiques visent à reconstruire l'histoire évolutive des organismes, des espèces et d'autres unités taxonomiques opérationnelles, une histoire considérée comme étant unitaire et permettant l'érection de systèmes classificatoires hiérarchiques. Ce qui est nécessaire pour reconstruire des systèmes phylogénétiques, c'est le lien similarité-phylogénie efficacement traduit par la notion d'homologie. Il faut donc très certainement comprendre le rôle de l'homologie, en tant que concept, pour comprendre la réalisation ou non des objectifs de la discipline. Cependant, ce concept traduit la croyance bien ancrée selon laquelle il est possible d'inférer des liens phylogénétiques sur la base de liens de similarité. La formation d'une catégorie délimitée par de tels liens est accessoire au processus

épistémique de la phylogénétique; en conséquence, l'utilisation épistémologique de la notion d'espèce naturelle est non justifiée.

L'ensemble de mon argument visant à répondre à la suggestion de Brigandt (2009) selon laquelle l'étiquette d'espèce naturelle GHP est utile pour comprendre la notion d'homologie peut être reformulé et résumé en ramenant le problème de la coordination dans le débat. Les espèces naturelles sont inaccessibles à l'observation directe, leur existence doit donc être inférée par l'entremise d'observations connexes. Les liens de similarité entre divers phénomènes et l'encadrement théorique de tels liens de similarité suffisent, selon Brigandt (2009) et la théorie GHP, à inférer l'existence d'espèces naturelles et à justifier l'utilisation de l'étiquette associée pour décrire des catégories comme celle d'homologie. Cependant, j'ai démontré que l'analyse épistémologique qui permet ultimement de postuler qu'une catégorie comme l'homologie est une espèce naturelle GHP *précède* un tel postulat. Ainsi, dire que l'analyse épistémologique de la similarité, des mécanismes qui sous-tendent cette similarité et des objectifs d'une discipline requiert une théorie des espèces naturelles est inexact. L'étiquetage des concepts scientifiques qui les associe au statut d'espèce naturelle découle de l'analyse épistémologique concernée et est donc tout à fait accessoire à cette analyse épistémologique; il s'agit d'un étiquetage *a posteriori*. La notion d'espèce naturelle est simplement utile si notre objectif est d'identifier des espèces naturelles. Si notre objectif est, au contraire, d'offrir des analyses détaillées d'un processus inférentiel, la notion est alors accessoire, voire trompeuse (puisque'elle suggère un point focal pour les analyses qui demeure injustifié; je reviens sur ce point en 4.3.3). Il est utile, à ce point-ci, de se rappeler que Conix et Chi (2020) reconnaissent aussi l'inutilité théorique de la notion d'espèce naturelle (mais suggèrent de la conserver pour d'autres raisons).

Revenons maintenant à l'arche des espèces naturelles. J'ai démontré au chapitre précédent que la légitimité des théories des espèces naturelles repose sur deux piliers, soit la croyance selon laquelle certaines catégories scientifiques ont un statut épistémique privilégié ainsi qu'un statut métaphysique privilégié. J'ai aussi insisté sur l'ancrage en esse qui solidifie la structure dans son ensemble. L'esse est le lien logique qui unit les deux piliers : la métaphysique explique l'épistémique ou l'épistémique permet d'inférer la métaphysique. Brigandt (2009) est un exemple paradigmatique de la faiblesse structurelle qui résulte de la ténuité de l'esse. Il

affirme explicitement que l'aspect métaphysique de l'homologie en tant qu'espèce naturelle explique son utilité épistémique, mais cet aspect métaphysique doit être accepté ou inféré sur la base de l'utilité épistémique de l'homologie puisque les espèces naturelles sont inaccessibles à l'observation directe. Il s'ensuit une circularité typique de l'arche des espèces naturelles : le statut métaphysique est déduit du statut épistémique qui est expliqué par le statut métaphysique. Ce cercle est utilisé pour fonder l'injonction à développer des théories des espèces naturelles.

Or, le cohérentisme historique nous apprend que la circularité d'un argument isolé est en soi normale et non problématique *si celui-ci contribue à satisfaire l'impératif du progrès*. J'ai démontré, dans l'analyse précédente, que les arguments de Brigandt (2009) visant à établir l'utilité épistémologique de la notion d'espèce naturelle GHP sont non satisfaisants et qu'on peut douter du fait que cette notion contribue au progrès de l'épistémologie de la classification. De plus, j'ai formulé mes contre-arguments de sorte qu'ils puissent être mobilisés même si l'utilité épistémologique des espèces naturelles est défendue sur la base d'une autre théorie que celle de Boyd. Bien qu'un argument plus exhaustif reste à faire en ce sens, mon argument éliminativiste suggère que les limites de l'utilité épistémologique de la notion d'espèce naturelle dépassent les limites de la théorie GHP.

4.3.3 Deux arguments auxiliaires au rejet de l'utilité épistémologique de la notion d'espèce naturelle

Au sein de la tradition des espèces naturelles, l'utilité épistémologique de la notion d'espèce naturelle est rarement défendue comme c'est le cas dans le texte de Brigandt de 2009 (je n'ai pu trouver d'autres défenses explicites du même genre). Or, considérant les faiblesses de l'argument tout juste exposées, je considère qu'il serait judicieux d'éviter de structurer l'épistémologie de la classification autour de la notion d'espèce naturelle. Cette prescription est soutenue par deux arguments auxiliaires : premièrement, les théories des espèces naturelles gagnent à être considérées comme des éléments internes aux pratiques classificatoires et plus spécifiquement des mécanismes de légitimation de ces pratiques; deuxièmement, elles tendent à focaliser l'analyse des pratiques classificatoires sur les catégories scientifiques et ainsi à négliger

le reste de la toile épistémologique impliquée. Dans les lignes qui suivent, je présente ces deux arguments tour à tour.

L'adhésion des théoriciens et des théoriciennes des espèces naturelles à l'arche décrite au chapitre précédent a des conséquences importantes sur leur façon d'appréhender les pratiques classificatoires. Cela mène notamment à une confusion entre outils et objets d'analyse : les théories des espèces naturelles utilisent l'idée même que certaines catégories détiennent un statut particulier (celui d'espèce naturelle) pour analyser des pratiques classificatoires en sciences; or, les praticiens et praticiennes considèrent, entre autres choses, que certaines catégories sont particulièrement utiles, réelles, etc., c'est-à-dire qu'elles sont des espèces naturelles (bien que cette expression ne soit pas toujours utilisée). Ce chevauchement des croyances inscrites au sein de l'arche des espèces naturelles et des croyances véhiculées par ce qui devrait être son objet d'étude fait de la tradition des espèces naturelles une approche mal adaptée à l'épistémologie des pratiques classificatoires.

Pour illustrer la chose, revenons à l'exemple de l'homologie, considérée comme une espèce naturelle (Assis et Brigandt 2009; Brigandt 2009; Pearson 2019; Rieppel 2015)²⁷. La citation suivante illustre bien la chose :

Homologies are unique evolutionary lineages of traits, but they are routinely evolutionarily *divergent* lineages. This undercuts their actually being individuals, but it does not prohibit conceiving of, and acting as though they are individuals. Homologies, therefore, really are natural kinds. (Pearson 2019, 42)

Dans ce passage qui conclut son article, Pearson affirme que l'homologie est une espèce naturelle (GHP) parce que seule une telle description métaphysique de l'homologie permet de rendre compte d'un fait qu'il considère comme étant indéniable (à l'instar de nombre de phylogénéticiens et de phylogénéticiennes), soit celui de la divergence des lignées impliquées au sein de certains processus évolutifs. Cette description métaphysique de l'homologie est donc choisie non pas pour son utilité épistémologique, mais bien pour sa cohérence avec des

²⁷ *Mutatis mutandis*, mon approche permettrait aussi d'analyser les arguments selon lesquels les homologies sont des individus historiques plutôt que des espèces naturelles puisque dans les deux cas, il s'agit d'une croyance entretenue par les scientifiques qui devrait être étudiée épistémologiquement et non utilisée comme axe central d'une épistémologie de la classification.

connaissances développées par les biologistes au travers du temps. La théorie GHP des espèces naturelles est ainsi un mécanisme qui offre, *a posteriori*, une cohésion accrue à l'appareillage de la phylogénétique par l'entremise d'une interprétation philosophique de la notion d'homologie qui concorde avec la théorie.

Les théories des espèces naturelles, comme la théorie des groupes homéostatiques de propriétés mobilisée par Brigandt ou Pearson pour rendre compte de la notion d'homologie, sont donc mieux comprises comme des éléments internes aux pratiques classificatoires²⁸. Elles font partie des stratégies mobilisées par les chercheurs et chercheuses pour légitimer des pratiques de sorte à dépasser les limites imposées par le problème de la coordination (elles contribuent à légitimer des procédures qui rendent observable l'inobservable, à solidifier les itérations d'un complexe discursif). En faisant de l'homologie une espèce *naturelle*, on renforce l'idée selon laquelle l'interprétation néo-darwinienne des liens de similarité entre états de caractères est apte à rendre compte de l'ordre *naturel*. En agissant de la sorte, on confère une légitimité aux inférences similarité-phylogénie qui dépasse largement ce qui est observable de façon directe, mais qui jouit du *respect* (Chang 2004) que les praticiens et praticiennes entretiennent à la fois à l'égard des pratiques préalables de leur champ disciplinaire et de la vénérable notion d'espèce naturelle.

Le rapport philosophique que Pearson (parmi d'autres) entretient avec les pratiques classificatoires est en ce sens un rapport de légitimation et non d'analyse. Les théories des espèces naturelles offrent un soutien métaphysique aux pratiques en décrivant l'ontologie des catégories scientifiques. Tout se passe ici en accord avec ce que van Fraassen appelle la perspective surplombante (voir chapitre 2). Cette perspective suggère d'adopter un regard anhistorique face aux pratiques scientifiques et de les interpréter à la lumière d'une théorie

²⁸ La tradition des espèces naturelles revient ainsi à ce que Hacking appelle un mécanisme de stabilisation. Il s'agit d'un outil rhétorique, institutionnel ou matériel qui confère à un style de raisonnement (les pratiques classificatoires sont pour Hacking et Crombie un style de raisonnement scientifique) une certaine stabilité. Cette stabilité est le fruit de contraintes sur la pratique qui contribuent aussi à solidifier le style de raisonnement face à des critiques externes. Dans le cadre de cette thèse, il m'est malheureusement impossible d'explorer le corpus de textes portant sur les styles de raisonnement scientifique en profondeur (voir notamment Crombie 1994, 1995 et Hacking 1994, 2009). Je tiens simplement à mentionner la notion de mécanisme de stabilisation qui a motivé et inspiré la présente critique de la tradition des espèces naturelles.

stabilisée en acceptant de fait la logique de celle-ci. Or, il a bien été démontré au chapitre 2 (grâce à van Fraassen, à Chang et à Tal) que cette perspective, prise seule, occulte des faits importants des systèmes épistémiques, notamment leur dynamisme, l'historicité des croyances et les différents facteurs qui expliquent tous ces changements. C'est de l'évolution de tels systèmes dont l'épistémologie doit faire son objet d'étude (Chang 2004; Tal 2016; van Fraassen 2008) et non du statut métaphysique et anhistorique de catégories isolées. Minimale, il faudrait séparer ces deux tâches plus clairement : les théories des espèces naturelles sont mal adaptées à l'étude des systèmes épistémiques complexes, soit la première tâche; en remplissant la deuxième, elles en viennent à faire partie de l'objet d'étude de l'épistémologie de la classification en tant que mécanisme de légitimation des pratiques classificatoires.

Le deuxième argument auxiliaire consiste à mettre en lumière les éléments des pratiques classificatoires qui tendent à être occultés par une approche épistémologique centrée sur les catégories scientifiques (plutôt que sur l'ensemble du système épistémique). Cet argument va plus loin que ce qui précède. J'ai jusqu'à présent argumenté que les théories des espèces naturelles sont inadéquates dans le cadre de mon projet épistémologique. Ici, j'argumente que l'épistémologie de la classification doit éviter de focaliser son propos sur les catégories scientifiques. La meilleure façon de défendre cette thèse est de souligner que les théories des espèces naturelles (par exemple, Brigandt 2009; Pearson 2019) ou d'autres approches similairement centrées sur les catégories scientifiques (comme Kendig 2016) rendent compte de l'homologie en tant que catégorie *sans appréhender la construction de systèmes phylogénétiques qu'elle rend possible*. Cette erreur est, selon moi, impardonnable (épistémologiquement parlant) puisqu'elle donne l'impression que l'objectif des praticiens et des praticiennes est de trouver des homologies, alors qu'il s'agit plutôt de reconstruire des systèmes phylogénétiques, de placer des nouvelles unités taxonomiques au sein de systèmes déjà disponibles, de réviser des systèmes à l'aune de nouvelles données, etc. L'épistémologie de la classification gagne ainsi à adopter une perspective plus englobante sur son objet d'étude plutôt que de se limiter à l'étude des catégories scientifiques et à leur statut métaphysique. La vision synoptique proposée par van Fraassen (voir chapitre 2) se prête bien à l'exercice, comme je l'illustre succinctement dans ce qui suit (la meilleure illustration de la valeur de cette perspective offerte dans le présent ouvrage demeure

celle présente aux chapitres 6 et 7; j’y adopte la perspective synoptique pour l’analyse de pratiques phylogénétiques contemporaines).

Par l’adoption d’une perspective surplombante, on comprend que le rôle inférentiel crucial occupé par l’homologie est, au sein de la phylogénétique traditionnelle, directement lié au fait que le cadre théorique dominant pose la divergence comme une tendance dominante de l’évolution des lignées biologiques (voir la citation de Pearson donnée plus tôt). Celles-ci tendent à diverger toujours plus les unes des autres, de sorte que seules les relations de similarité qui traduisent cette divergence peuvent servir de matière inférentielle. Ainsi, la croyance selon laquelle le patron arborescent (qui sous-tend la divergence) est le meilleur modèle pour rendre compte de l’évolution précède l’utilisation phylogénétique de l’homologie, du moins logiquement (l’homologie est utile *parce que* la similarité due à des phénomènes comme la convergence est jugée inapte à soutenir des inférences phylogénétiques). Elle la précède, puis lui succède, considérant que les reconstructions phylogénétiques peuvent être utilisées pour établir des systèmes taxonomiques hiérarchiques dont la légitimité provient elle aussi du fait que ces systèmes calquent le patron divergent.

Dans une perspective interne au processus historique en jeu, la relation entre ces croyances diverses est tout autre. Les systèmes hiérarchiques trouvent leur source dans les méthodes d’analyses comme celle associée à l’Arbre de Porphyre (voir chapitre 3). Le tout en est venu à structurer la biologie comparative et la taxonomie, de sorte que Linné (et d’autres naturalistes de l’époque) a pu calcifier l’utilisation de ces systèmes hiérarchiques et de la nomenclature qui l’accompagne. Ce processus a permis l’émergence graduelle du principe d’homologie, d’abord implicite dans les idées ancestrales de « similarité malgré la différence » (Kendig 2016; Nelson 1994; Rieppel 1994), puis explicite dans la distinction entre analogues et homologues (Owen 1843). Le tout est finalement passé du statut de méthode de classification utile (Stevens 1997) à celui de description de l’ordre naturel, une transition s’étant cristallisée avec la théorie darwinienne et la métaphore arborescente (Panchen 1992), de nos jours considérée comme condition première de la systématique phylogénétique (Felsenstein 2004; O’Hara 1997; Wiley et Lieberman 2011).

Cette analyse succincte permet de comprendre que l'homologie réfère à beaucoup plus que des groupes de propriétés homéostatiques. C'est un axe de pivot sur lequel s'est construite la trajectoire évolutive de la biologie comparative, une croyance qui s'appuie sur bien d'autres et qui en soutient encore plus, un élément historique dont les relations se sont modifiées au cours du temps. Les approches centrées sur l'idée que l'homologie est une catégorie positive ou naturelle négligent, par leur conception anhistorique de l'objet d'étude, toutes ces dimensions historiques dans le but implicite de légitimer un ensemble de pratiques, ou du moins, d'en donner une interprétation philosophique. En ce sens, bien que l'analyse de catégories isolées soit potentiellement une source d'informations utiles (l'approche synoptique suggère de croiser le regard surplombant et celui interne aux processus historiques), elle est visiblement insuffisante pour rendre compte des pratiques classificatoires et de leur dynamisme inhérent. Elle fait abstraction de données nécessaires à la bonne compréhension d'un objet d'étude comme les pratiques classificatoires de la biologie de l'évolution. Il convient de noter qu'il y a probablement plus d'éléments épistémiques dans l'angle mort des approches centrées sur les catégories que ceux mentionnés ici (par exemple, les croyances monistes des praticiens et praticiennes, l'apport des technologies sur les pratiques d'observation et de description, les principes d'analyse de la similarité, etc.), mais ceux-ci suffisent pour l'instant à évoquer l'ampleur de cet angle mort.

4.4 Conclusion

La prescription la plus importante de l'éliminativisme à l'égard des espèces naturelles est la suivante : l'étude philosophique des pratiques classificatoires en sciences devrait être structurée autrement que par l'entremise de ce concept. Pour Hacking (2007a) et Ludwig (2018), les théories spécifiques ayant été développées au sein de la tradition des espèces naturelles sont utiles pour rendre compte de certaines pratiques classificatoires, mais aucune théorie n'est apte à rendre compte de l'ensemble de ces pratiques. Il serait donc nuisible de s'entêter sur la voie du monisme théorique suggéré par un concept général d'espèce naturelle. Cette critique s'applique aussi aux approches comme celles de Boyd, de Dupré ou de Slater qui sont pluralistes quant à la teneur ontologique des espèces naturelles, mais monistes quant à leur statut épistémique. Brigandt, quant à lui, critique surtout l'utilisation du prédicat « naturelles » pour parler des espèces, une utilisation non justifiée qui présente des risques épistémiques et sociaux en

attribuant une immuabilité aux catégories ainsi décrites. Cette critique de Brigandt est doublée d'une proposition selon laquelle les catégories scientifiques gagnent à être analysées à l'aune des objectifs qu'elles permettent de remplir (ou non) plutôt que de leur statut métaphysique.

L'arche des espèces naturelles est ainsi attaquée sur deux fronts. Ludwig critique l'arc ou la croyance selon laquelle des théories générales des espèces naturelles pourraient contribuer à l'épistémologie des pratiques classificatoires. Brigandt remet en question la validité d'un des deux piliers, soit celui qui traite du statut métaphysique de *certaines* catégories scientifiques privilégiées. Au contraire, Conix et Chi (2021) ont formulé un argument afin de défendre l'utilisation du concept d'espèce naturelle. Cette notion aurait un rôle d'investigation et un rôle social qui justifient de la maintenir au cœur de l'épistémologie de la classification. À cet argument, j'ai répondu que l'inutilité théorique de la notion d'espèce naturelle (admise par Conix et Chi) rend problématique son utilisation pour remplir les deux rôles mentionnés.

Après avoir ainsi complété la revue du corpus de textes portant sur les espèces naturelles, j'ai étayé, dans la deuxième partie du chapitre, mon argument éliminativiste. Celui-ci suggère que la tradition des espèces naturelles est, dans le cadre d'une épistémologie des pratiques classificatoires en sciences, dotée d'une portée explicative limitée. Cela se comprend par le fait que cette tradition est structurée autour de l'arche des espèces naturelles. Au cœur des théories des espèces naturelles repose le constat de la légitimité épistémique des catégories scientifiques, une légitimité qui devrait être l'objet d'étude de l'épistémologie de la classification plutôt qu'un élément accepté *de facto*.

De ce constat suivent les trois points saillants de ma posture éliminativiste à l'égard des espèces naturelles : 1) le travail épistémologique traditionnellement attribué aux espèces naturelles (Brigandt 2009) peut être réalisé sans leur faire appel; 2) les théories des espèces naturelles sont inaptes à analyser de façon critique les pratiques classificatoires puisqu'elles font partie des mécanismes de justification et de légitimation des pratiques classificatoires; 3) les approches épistémologiques s'intéressant uniquement aux catégories scientifiques délaissent le reste de l'appareillage épistémique des pratiques classificatoires, par exemple les modèles qui

permettent de mettre en relation les catégories identifiées ou les croyances qui rendent l'observation des relations de similarités possibles.

Le tout est bien illustré par l'exemple de l'homologie. Brigandt (2009) argumente que pour bien comprendre son pouvoir inférentiel, il faut la considérer comme une espèce naturelle GHP. Or, une analyse détaillée d'un processus inférentiel typique de la phylogénétique témoigne du contraire : que l'homologie soit une espèce naturelle ou non, elle remplit son rôle parce qu'elle traduit le lien inférentiel entre similarité et phylogénie qui permet de discriminer les relations de similarité pertinentes de celles qui ne le sont pas. *Ce n'est pas l'existence d'une espèce naturelle qui justifie la validité de telles inférences, mais bien le fait qu'elles soient ancrées dans un système de croyances englobant qui a, par le passé, démontré son efficacité.* En d'autres termes, les théories des espèces naturelles considèrent qu'un fait concernant les catégories scientifiques explique le pourquoi et le comment des pratiques classificatoires. Au contraire, j'insiste sur le fait selon lequel c'est l'ensemble des relations entre les croyances composant un système épistémique qui explique les modalités classificatoires. De ce fait, je reste loin des théories des espèces naturelles afin d'analyser les pratiques phylogénétiques contemporaines. Bien que mon argument leur soit complémentaire, il convient de noter qu'il va plus loin que celui de Ludwig (2019) qui suggère d'abandonner les théories générales des espèces naturelles ou celui de Brigandt (2020) qui suggère d'abandonner l'idée de naturalité.

Chapitre 5 – L’homologie et la transition moléculaire

Once, Damaya would have protested the unfairness of such judgments. The children of the Fulcrum are all different: different ages, different colors, different shapes. Some speak Sanze-mat with different accents, having originated from different parts of the world. One girl has sharp teeth because it is her race’s custom to file them; another boy has no penis, though he stuffs a sock into his underwear after every shower; another has rarely had regular meals and wolfs down every one like she’s still starving. [...] One cannot reasonably expect sameness out of so much difference, and it makes no sense for Damaya to be judged by the behavior of children who share nothing save the curse of orogeny with her.

- Nora Keita Jemisin, dans *The Fifth Season*

5.1 Introduction

L’objectif principal de ce chapitre est de présenter la transition vers les données moléculaires qui caractérise la biologie de la deuxième moitié du 20^e siècle et tout particulièrement l’impact de cette transition sur la phylogénétique. La montée en puissance des données moléculaires est cruciale pour comprendre les pratiques contemporaines au sein desquelles gènes et protéines occupent une place prépondérante. Je profite de l’occasion pour remplir un objectif connexe, soit de démontrer l’utilité des analyses structurées par le modèle hélicoïdal du cohérentisme historique en contraste avec les analyses structurées sur l’arche des espèces naturelles. Mon analyse permet en effet d’identifier une différence importante entre l’homologie moléculaire et l’homologie morphologique. Bien que les deux soient définies similairement, leur mise en pratique se distingue fortement puisque l’analyse de similarité associée à l’homologie moléculaire requiert des hypothèses quant aux transformations des séquences génétiques alors que l’analyse de la similarité morphologique ne requiert rien de tel. Ainsi, le cohérentisme historique, en insistant sur l’analyse des pratiques permettant de dépasser les contraintes associées à la coordination entre théorie et phénomènes, fait apparaître de tels éléments constitutifs des pratiques scientifiques qui restent occultés par d’autres modèles épistémologiques, comme l’arche des espèces naturelles.

Il convient de préciser que j’aurais pu utiliser, dans le cadre de ce chapitre, d’autres catégories scientifiques que l’homologie comme point focal d’analyse. Au sein de la

phylogénétique, d'autres catégories sont candidates au statut d'espèce naturelle : les espèces biologiques et autres rangs taxonomiques, les familles de gènes, les gènes, les protéines, etc. Puisque l'homologie traduit le rapport entre similarité et phylogénie, elle occupe cependant une place de choix au sein de la structure inférentielle de la phylogénétique. Son analyse permet donc d'illustrer adéquatement la capacité du cohérentisme historique à appréhender des systèmes inférentiels complexes et de contraster efficacement la chose avec les apports de la tradition des espèces naturelles. De surcroît, les débats entourant l'homologie font écho à ceux portant sur la notion d'espèce biologique (Rieppel 2015; Pearson 2019), mais permettent de mieux prendre le pouls de la transition vers les données moléculaires et son impact sur la phylogénétique. Bien que je ne puisse effectuer l'exercice en détail, tout porte à croire que les conclusions ici défendues pourraient être transférées à cette dernière notion ainsi qu'à toute analyse de la phylogénétique axée sur de potentielles espèces naturelles.

Afin de réaliser ces objectifs, le chapitre est divisé en quatre parties. Dans la première partie, je décris, dans une perspective historique, la transition au moléculaire et la transformation de la phylogénétique qui l'a accompagnée. Les phylogénétiques moléculaire et morphologique sont présentées comme des itérations distinctes de la systématique (bien que similaires en de nombreux points). Dans la deuxième partie, je m'intéresse tout particulièrement à la notion d'homologie telle qu'elle est mobilisée dans le cadre de la biologie moléculaire. La troisième partie revient sur le traitement que la tradition des espèces naturelles a réservé à la notion d'homologie, notamment pour en noter les angles morts. La dernière partie analyse l'homologie au sein des deux itérations de la phylogénétique concernées (morphologique et moléculaire) afin d'explorer les transformations subies par le processus inférentiel de la discipline.

5.2 La transition moléculaire

La découverte de la structure de la molécule d'ADN, en 1953, par Franklin, Gosling, Watson et Crick confirme la croyance déjà bien ancrée, en génétique, selon laquelle l'hérédité organismique aurait un support matériel présent dans les cellules. En offrant une caractérisation physique précise de ce support matériel, les chercheurs et chercheuses impliqués dans cette découverte ont ouvert un monde de possibilités aux biologistes de leur époque. La génétique,

jusqu'alors une science n'ayant eu d'autre choix que de reposer sur une notion abstraite du gène, se voyait ainsi offrir un ancrage matériel, une cible à observer, manipuler et au sujet de laquelle des spéculations extrêmement fertiles allaient suivre (Le Guyader 2001).

L'influence de cette découverte sur les pratiques de classification en biologie est particulièrement importante. La biologie comparative avait été, jusqu'alors, principalement tournée vers l'analyse de la similarité morphologique et des signaux phylogénétiques qui peuvent en être issus. L'un des défauts évidents des méthodes concernées est leur faible potentiel de généralisation : des analyses portant sur les ailes d'animaux variés, par exemple, sont difficilement comparables à l'étude des flagelles bactériennes ou des spores fongiques. Ainsi, les connaissances phylogénétiques acquises par l'analyse morphologique ont mené principalement à la reconstruction d'arbres localisés, délaissant l'ambition darwinienne de reconstruire un arbre regroupant l'ensemble du monde vivant (O'Malley 2016).

La promesse accompagnant la découverte et l'exploration des données moléculaires, au contraire, était celle d'une abondance de données permettant de comparer tous les êtres vivants, de rendre comparable ce qui était auparavant incomparable. La transition vers les données moléculaires a nivelé les différences irréconciliables au sein du monde vivant de sorte à mettre en relation tous les clades qui le constituent. La légitimité de cette transition a d'abord été établie par la congruence entre les arbres phylogénétiques que les données moléculaires permettent de reconstruire et ceux issus des données morphologiques (O'Malley 2016; Suárez-Díaz 2014; Zuckerkandl 1987).

Malgré cette congruence initiale, le tournant du 21^e siècle a vu les données morphologiques perdre du terrain au profit de l'information issue du monde moléculaire. Cette transition a été facilitée, voire encouragée, par le développement de la phénétique à partir des années 1960 (Hull 2001; Panchen 1992; Sneath et Sokal 1973). Approche de la classification biologique aujourd'hui déchuë, la phénétique visait à classer les organismes en toute neutralité théorique, objectif absurde au regard de l'épistémologie informée par le problème de la coordination et la charge théorique de toute observation. Malgré sa posture philosophique douteuse, la phénétique a contribué de manière durable au projet phylogénétique en établissant

et en raffinant l'analyse quantitative de la similarité. Il convient de noter que l'expression « phénétiqne » est parfois utilisée pour référer aux pratiques d'analyse numérique de la similarité sans évoquer le courant de pensée qui est à l'origine de ces pratiques.

En effet, les méthodes d'analyse quantitatives ont rapidement été intégrées à l'appareillage de la cladistique (Farris 1970; Kluge et Farris 1969; Panchen 1992; Wiley 1981). La teneur de cette transformation de la cladistique se trouve notamment dans le calcul des longueurs de branches reliant les unités taxonomiques opérationnelles. Les états de caractères de ces unités taxonomiques sont d'abord polarisés, comme la cladistique hennigienne le recommandait déjà (c'est ce qu'on appelle la polarisation des caractères). Ainsi, les apomorphies (états de caractères dérivés) peuvent être distinguées des plésiomorphies (états de caractères hérités d'ancêtres plus lointains). Les apomorphies sont ensuite utilisées pour construire des réseaux wagnériens (Farris 1970; Wiley 1981; Wiley et Lieberman 2011) ou reconstruire des arbres phylogénétiques plus directement en postulant les états de caractère d'ancêtres non connus des unités taxonomiques étudiées (Kluge et Farris 1969). Ces deux méthodes permettent ultimement de calculer les distances évolutives entre les unités taxonomiques et ainsi de choisir la topologie qui minimise le nombre d'événements évolutifs (spéciations, transformation d'un état de caractère en un autre, etc.).

La combinaison d'un nouveau type de données permettant la comparaison pour l'ensemble du monde vivant, de l'augmentation de la puissance calculatoire issue de développements technologiques et du tournant quantitatif de la discipline a catapulté l'analyse phylogénétique au cœur de l'étude de la biologie : « Biology's incursions into molecules and genomes have transformed an otherwise stale and isolated discipline of evolutionary study into one of the most vibrant, central and important facets of our science. » (Woese 1998, 11044). Malgré toutes les promesses du tournant moléculaire, les limites de cette nouvelle perspective sont vite constatées. Pour le dire simplement, le traitement phylogénétique des données moléculaires a mené à des incongruences significatives à au moins deux niveaux d'analyse : d'abord, entre des phylogénies issues de données moléculaires; ensuite, entre des phylogénies issues d'analyses moléculaires et morphologiques (O'Malley 2016). De telles incongruences sont considérées comme étant problématiques par la communauté des chercheurs et chercheuses en

phylogénétique, celle-ci étant guidée par l'espoir de retracer l'histoire unique du monde vivant (Baptiste 2007; Doolittle et Baptiste 2007). Ultimement, ces deux types d'incongruence, loin de paralyser la communauté, ont stimulé les recherches axées sur la méthodologie afin de développer des procédures pour identifier les incongruences issues d'artéfacts et pour que le plus de données possible (même de types variés) puissent être analysées communément afin d'obtenir un regard aussi exhaustif que possible sur l'histoire des êtres vivants (Kearney et Rieppel 2006; O'Malley 2013, 2016).

Il a en effet d'abord été pensé que la présence d'incongruences résultait d'un manque de données disponibles, croyance qui persiste encore à ce jour, bien qu'elle ait, de tout temps, rencontré une résistance soutenue et articulée :

Before the flood of genome sequence data was fully underway, it was already argued that simply having 'more data would not produce better trees, because the models underlying inference methods are [clearly] not accommodating the processes that have shaped the data' (Naylor et Brown 1998, p. 71). (O'Malley 2016, 78-79)

Cette citation évoque une idée centrale à mon propos, soit le fait que la nature et la quantité des données utilisées ne peuvent mener à des succès scientifiques qu'en suivant les règles d'un jeu délimité par les modèles et les croyances qui structurent leur observation et leur traitement. En ce sens, la phylogénétique moléculaire, si elle mobilise un ensemble de croyances similaire à celui de la phylogénétique morphologique, ce qui est indéniablement le cas comme en témoigne la transition vers la taxonomie quantitative (phénétique) qui a été appliquée aux deux types de données, alors elle doit abandonner l'espoir de régler les problèmes fondamentaux auxquels fait face la systématique évolutionniste depuis ses débuts. Pour le dire autrement, la phylogénétique moléculaire est une itération de la biologie comparative qui s'inscrit dans un rapport intime avec les itérations précédentes sans s'en écarter drastiquement. Utiliser la notion d'homologie comme point focal de l'analyse épistémologique de cette succession d'itérations épistémiques permet de prendre le pouls de certaines des différences et similarités qui les caractérisent.

5.3 L'homologie moléculaire

L'homologie, rappelons-le, se définit à un haut niveau d'abstraction comme étant la similarité considérée comme théoriquement pertinente à la lumière d'une théorie (Pavlinov 2012). De façon un peu plus précise (sous le joug du néo-darwinisme), elle réfère à la similarité résultant de l'ascendance commune de deux unités taxonomiques opérationnelles (Hall 2007; Kapli, Yang et Telford 2020). En ce qui concerne son opérationnalisation, certains critères sont utilisés pour analyser la similarité morphologique de sorte à formuler des hypothèses d'homologie : structure des traits observés, présence d'intermédiaires entre deux états de caractères distincts et position des traits par rapport au reste de l'organisme qui les porte (Remane 1952). Or, l'importance à accorder à ces critères, la validité de ces critères et leurs poids respectifs face aux phylogénies reconstruites ou préalables sont tout autant de questions qui ont contribué à maintenir en vie le « problème » de l'homologie, un problème relatif à son opérationnalisation plutôt qu'à sa définition (de Beer 1971; Inkpen et Doolittle 2016). Au premier regard, la transition moléculaire suggère une simplification de la mise en œuvre de l'homologie :

Molecular phylogeneticists inherited the concept of homology from their morphological counterparts, and as such relied on similarity as well, just of a molecular rather than morphological type (Patterson 1987). At the time (early 1960s), there was a sense among many that a molecular-based homology concept would be less problematic, or at least cleaner, than a morphological one, as amply illustrated by the various contributors to Bryson et Vogel's (1965) epochal *Evolving Genes and Proteins*. And this was directly linked to the problem of trustworthy similarity. (Inkpen et Doolittle 2016, 187)

L'idée sous-jacente était que les données moléculaires permettraient de développer une métrique univoque pour quantifier la similarité. Plus précisément, les nucléotides (ou acides aminés) qui forment deux molécules à l'étude peuvent être comparés, mis côte à côte, de sorte à établir lesquels se correspondent (je reviens sur le détail des méthodes d'alignement dans la quatrième partie du présent chapitre). Dans cette optique, le fait que deux séquences présentent le même nucléotide au même endroit implique une relation de similarité entre les deux séquences, de sorte que l'alignement de séquences est rapidement devenu une manière d'envisager l'homologie au niveau moléculaire : « DNA is complex enough to preclude analogy as a cause of overall similarity [... thus,] we finally have a method that can sort homology from

analogy. » (Gould 1985, cité dans Inkpen et Doolittle 2016, 187). La complexité à laquelle Gould réfère est issue de la longueur des séquences génétiques comparées doublée de l'existence de quatre types de bases azotées.

La simplification conceptuelle et pratique de l'homologie suggérée par la transition moléculaire a cependant laissé émerger de nouvelles difficultés rappelant de nombreux problèmes déjà observés avec l'opérationnalisation de l'homologie morphologique : « Molecular phylogeneticists sometimes talk as if their discipline has resolved these controversies: it has not. » (Inkpen et Doolittle 2016, 184). La similarité, même au niveau moléculaire, est équivoque, surtout lorsqu'elle est mobilisée pour inférer l'histoire des molécules concernées ou des organismes qui les portent. Voici trois exemples de complications liées à l'étude du lien inférentiel liant similarité moléculaire et phylogénie.

Premièrement, il est indéniable que la similarité moléculaire qui caractérise deux organismes ou unités taxonomiques peut être le fruit de différents processus évolutifs qui ont des conséquences bien distinctes sur notre capacité à utiliser cette similarité de sorte à inférer des trajectoires phylogénétiques (Inkpen et Doolittle 2016). Plus spécifiquement, la divergence qui caractérise deux séquences peut être le fruit d'un événement de spéciation, de duplication ou de transferts génétiques latéraux (ou horizontaux); les homologues sont alors appelés, respectivement, des orthologues, des paralogues et des xénologues (ce qui rend la synonymie entre homologie et synapomorphie insoutenable). Ce sont principalement les orthologues qui sont utilisés pour retracer les phylogénies des unités taxonomiques (au niveau des organismes) qui les portent puisque leurs lignées sont calquées sur celles des organismes (par définition). Ainsi, les méthodes pour distinguer les différents types d'homologie sont nécessaires pour la phylogénétique moléculaire (Kapli, Yang et Telford 2020), au même titre que la cladistique a développé des outils pour distinguer synapomorphies et symplesiomorphies afin d'exclure les groupes paraphylétiques des classifications qu'elle cherche à produire.

Deuxièmement, et ce point découle du premier, force est d'admettre que l'histoire évolutive d'une espèce peut être distincte de celle des gènes qu'on utilise habituellement pour la retracer. Le fait de distinguer les orthologues des autres types d'homologues n'offre cependant

qu'une solution partielle à ce problème. En effet, même si l'étude des orthologues témoigne d'une succession d'embranchements potentiellement associée aux lignées d'organismes, ce processus de divergence ne résume qu'une partie de l'histoire de la lignée. La partie ainsi négligée est fonction de la partie du matériel génétique qui ne répond pas des mêmes mécanismes de transmission que les orthologues, et celle-ci peut être très significative, notamment dans le cas du monde microbien : « What does it mean, then, to speak of an organismal genealogy when nearly all of the genes in the cell, genes that give it its general character, do not share a common history? This question goes beyond the classical Darwinian context » (Woese 2002, cité dans Doolittle et Baptiste 2007, 2047). Cette fâcheuse situation, l'une des motivations principales pour la transition vers les réseaux phylogénétiques étudiée au chapitre 7 de cette thèse, témoigne d'une complication significative par rapport à la notion d'homologie en rapport à l'objectif de la phylogénétique. Le tout témoigne aussi de la dépendance des pratiques phylogénétiques traditionnelles à une conception arborescente de l'évolution du monde vivant, comme je le détaille au chapitre 6.

Une troisième complication relative à l'utilisation moléculaire de la notion d'homologie est un problème fondamental relatif à l'opérationnalisation du concept : « If indeed most (possibly all) genes are 'at some level homologous' and yet we persist in wanting to decide of any two genes whether they are 'homologous or not', then we must impose some further criteria. » (Inkpen et Doolittle 2016, 190). Ce qu'Inkpen et Doolittle soulignent par cette citation est le fait que les critères pour l'analyse de la similarité moléculaire sont tout aussi équivoques qu'au niveau morphologique. J'ai mentionné plus tôt que l'alignement de séquences représente une méthode largement utilisée pour l'analyse de la similarité génétique. Cependant, le critère de similarité structurelle des séquences que l'alignement incarne (il faut que les nucléotides ou acides aminés soient organisés d'une certaine manière pour que leur présence au sein de la séquence soit jugée théoriquement pertinente) n'est qu'une option parmi d'autres. Ce point est parfaitement illustré par le développement récent de méthodes de comparaison sans alignement (*alignment-free phylogenetics*). Ces méthodes, très diversifiées, peuvent être divisées en deux sous-groupes : les méthodes mesurant la fréquence de petits segments d'ADN au sein des séquences que l'on cherche à comparer et celles qui comparent le contenu informationnel des séquences dans leur

ensemble (Zielezinski et al. 2017; Vinga 2014). Dans les prochaines lignes, je présente rapidement la première de ces deux familles de méthodes simplement pour illustrer la teneur des méthodes de rechange à celles basées sur l'alignement.

L'analyse de la similarité entre deux séquences génétiques (qui peuvent être des génomes entiers) sans alignement peut être effectuée en comparant la fréquence de *mots* (ou *k-mers*, petits segments d'ADN ou d'ARN) au sein des séquences étudiées. Ces méthodes sont conceptuellement simples (Bernard et al. 2017; Zielezinski et al. 2017; 2019). Les séquences comparées sont d'abord divisées en mots, typiquement des sous-séquences de 3 à 25 nucléotides. Un ensemble de mots est ensuite construit, incluant seulement les mots qu'on retrouve au moins une fois dans chaque séquence comparée, puis on comptabilise le nombre d'occurrences de chacun des mots dans chacune des séquences. Les données ainsi comptabilisées, sous forme de vecteurs, sont soumises à une fonction qui permet de quantifier la relation de similarité liant les séquences. Cette similarité quantifiée peut finalement être utilisée pour inférer des relations phylogénétiques (Bernard et al. 2017; Zielezinski et al. 2017; 2019).

Au-delà du détail, il est important de saisir que ces méthodes permettent d'évaluer la similarité de deux séquences génétiques sans les aligner. Le critère de la similarité structurelle est ainsi délaissé au profit de l'observation de la fréquence de sous-segments. Cette famille de méthodes a de nombreux avantages (selon ses défenseurs et défenseuses) : la complexité des calculs augmente seulement de façon linéaire avec la longueur des séquences (et non de façon exponentielle, par exemple); de telles analyses sont moins susceptibles d'être induites en erreur par les événements de brassage ou de recombinaison génétique; elles sont plus aptes à reconstruire des phylogénies lorsque le signal phylogénétique est faible, par exemple lorsqu'il y a un taux très bas de conservation des séquences étudiées (Agüero-Chapin et al. 2016; Bernard et al. 2017; Noviello et al. 2018; Pham, Gao et Phan 2017; Ragan et Chan 2013; Vinga 2014; Zielezinski et al. 2017).

Malgré le fait que cette méthode d'analyse de la similarité génétique offre des résultats qui convergent avec les résultats de méthodes plus traditionnelles (les analyses basées sur les

données moléculaires avaient de la même façon acquis leur légitimité par leur capacité à reproduire des résultats issus de l'analyse morphologique), l'analyse sans alignement s'est frottée, du moins par le passé, à une résistance digne de mention (Vinga et Almeida 2003; Zieleszinski et al. 2017; 2019). Cette résistance fait notamment intervenir l'idée d'homologie. Selon ses détracteurs et détractrices, l'analyse sans alignement identifie une similarité non homologique. Pourtant, l'analyse sans alignement est *a priori* compatible avec la définition générale de l'homologie et *a posteriori*, ses succès empiriques devraient renforcer sa validité. Au-delà de la validité de cette posture et de la méthode concernée, il convient ici de lire le sous-texte : le « problème » de l'homologie, qui relève de l'opérationnalisation de cette notion, est loin d'avoir atteint son terme avec la transition moléculaire. Plusieurs critères d'analyses de similarité sont présentement disponibles et peuvent guider les inférences phylogénétiques.

Ces trois enjeux relatifs à l'opérationnalisation de l'homologie moléculaire (et d'autres encore; Haggerty et al. 2014; Inkpen et Doolittle 2016; Kapli, Yang et Telford 2020; Morrison, Morgan et Kelchner 2015), bien que critiques pour la discipline à l'étude puisqu'il s'agit d'éléments qui influent tout le processus inférentiel impliqué, ne menacent aucunement le sens général attribué à la notion d'homologie dans le contexte phylogénétique : deux traits sont homologues si la similarité qui les lie témoigne d'une relation phylogénétique bien précise, soit l'ascendance commune. Ainsi, bien que les critères de similarité utilisés et la matière étudiée soient différents de ceux de la phylogénétique morphologique, la notion conserve un sens qui incarne l'appartenance *sine qua non* de la phylogénétique arborescente au paradigme néodarwinien.

5.4 De l'analyse épistémologique de l'homologie en termes d'espèce naturelle

Au sein de la tradition des espèces naturelles, il est frappant de constater que la plupart des écrits portant sur la notion d'homologie l'appréhendent au niveau morphologique, que ce soit dans une perspective phylogénétique ou développementale. Des écrits influents comme ceux de Brigandt (2009), de Griffith (2007), de Pearson (2019) ou de Kendig (2016) évoquent rapidement l'aspect moléculaire de la chose avant de centrer leur propos sur le niveau

d'organisation plus élevé qu'est celui de l'organisme, de ses traits et des modules développementaux qui les sous-tendent. D'un point de vue épistémologique, l'homologie moléculaire gagne donc à être considérée plus directement, notamment parce que les données génétiques et protéiques sont de nos jours incontournables en phylogénétique (O'Malley 2016). Avant d'utiliser le cohérentisme historique pour analyser le rapport entre les itérations morphologiques et moléculaires de la phylogénétique, il convient de synthétiser mon argument éliminativiste à l'égard des espèces naturelles, en plus de démontrer la pertinence de cette critique relativement à l'homologie et à son incorporation au sein des différentes itérations phylogénétiques.

Pour Owen (1843), deux traits sont homologues s'il s'agit de deux occurrences du *même* trait sous différentes formes et fonctions. La question pérenne, depuis cette époque et même avant, est de savoir ce que le terme « même » implique (de Beer 1971; Inkpen et Doolittle 2016; Morrison, Morgan et Kelchner 2015; Nelson 1994; Panchen 1994). Pour Owen, cela implique une relation des traits des organismes avec ce qu'il appelle l'archétype, une notion à saveur platonicienne. Pour Darwin, la définition de l'homologie est virtuellement la même, la notion d'ancêtre commun remplaçant celle d'archétype. Cette substitution définitionnelle a cependant guidé la biologie comparative sur une trajectoire de sophistication des outils phylogénétiques, comme décrit au chapitre premier de cette thèse. Cette sophistication a engendré un morcellement de l'homologie en concepts plus spécifiques, autant pour l'analyse morphologique (symplesiomorphies et synapomorphies) que pour l'analyse moléculaire (orthologues, paralogues, xénologues). Nonobstant cette radiation du concept (Brigandt 2003), les notions plus spécifiques renvoient toutes, d'une manière ou d'une autre, à l'idée clé de voûte selon laquelle la relation d'homologie (et celles plus précises qu'elle englobe) est un lien de similarité qui témoigne d'une ascendance commune : « a working definition of homology is the presence of the same feature in two [or more] organisms whose most recent common ancestor also possessed the feature » (Hall 2007, 379, cité dans Morrison, Morgan et Kelchner 2015, 47).

Depuis la transition darwinienne, d'un point de vue strictement théorique ou définitionnel, la notion a joui d'une impressionnante stabilité. En cohérence avec ce constat, la déclaration selon laquelle l'homologie est une espèce naturelle GHP (Assis et Brigandt 2009;

Brigandt 2009; Pearson 2019; Rieppel 2015) convient tant aux occurrences moléculaires que morphologiques de cette relation; cette déclaration s'arrime autant aux xénologues et aux paralogues qu'aux synapomorphies et aux simplésiomorphies, sans oublier les homologies transformationnelles et taxiques. En effet, dans toutes ces situations, les traits dits homologues se ressemblent parce que des mécanismes évolutifs (notamment l'hérédité), développementaux (pour les homologues morphologiques) et spécifiquement génétiques (comme les mécanismes de réparation de l'ADN) viennent soutenir la coréalisation des propriétés qui caractérisent les homologues. Cette uniformité définitionnelle explique peut-être pourquoi nombre de théoriciens et théoriciennes des espèces naturelles ont négligé l'étude de l'homologie moléculaire qui semble, au premier regard, n'être qu'une simple adaptation de l'homologie morphologique.

Parler de l'homologie comme étant une espèce naturelle GHP, c'est insister sur le potentiel unificateur de la dimension théorique de l'homologie. Ce geste est toutefois posé au détriment de l'étude des contextes inférentiels plus spécifiques dans lesquels elle est utilisée (voir chapitre précédent). En démontrant, grâce à la théorie GHP, que les homologies développementale et phylogénétique (par exemple) sont compatibles (Assis et Brigandt 2009; Brigandt 2009; Pearson 2019; Rieppel 2015), la tradition des espèces naturelle détourne notre regard des différences patentes qui qualifient l'utilisation de l'homologie au sein des diverses sous-disciplines biologiques au profit d'un ancrage *naturel* et univoque de la notion. De surcroît, l'approche en termes d'espèces naturelles implique une perspective surplombante par rapport à l'objet d'étude (van Fraassen 2008; voir chapitre 2 de cette thèse) puisqu'une lecture historique du même phénomène témoigne de la construction progressive du concept, ce qui mine sa *naturalité*. Cette vision de l'homologie comme concept historique est donc tue au profit du point de vue surplombant, c'est-à-dire du point de vue de la théorie néo-darwinienne stabilisée. Dans cette perspective, l'homologie rend *de fait* possible la coordination des aspects théoriques de la biologie de l'évolution avec l'empirie; il s'agit d'une relation considérée comme étant réelle indépendamment de nos pratiques épistémiques : « Homology exists independently of our ability to recognize it. Indeed, both homologies and phylogenies need to be 'discovered' within the phenotypic and genotypic data that we have accumulated about biological organisms. » (Morrison, Morgan et Kelchner 2015, 55).

Le point de vue « naturaliste » (Ereshefsky et Reydon 2015; Khalidi 2013) qu'adopte la tradition des espèces naturelles se fond avec la perspective surplombante en question et avec le point de vue des praticiens et praticiennes sur la question (comme en témoigne la citation précédente). En effet, j'ai démontré que l'arche des espèces naturelles embrasse la validité de catégories scientifiques comme l'homologie et cherche à en démontrer la naturalité. L'arche est en ce sens un modèle qui renforce la légitimité de certaines catégories scientifiques. Les approches GHP de l'homologie en sont un bon exemple; elles renforcent la légitimité de la lecture réaliste de l'homologie (Brigandt 2002; Pavlinov 2012) en lui offrant un ancrage philosophique. Cet ancrage est toutefois trivial : la théorie GHP nous dit que les homologies, définies comme des relations de similarités expliquées par l'existence de mécanismes évolutifs (hérédité, formation de lignées, divergence, fluidité génétique au sein des populations, etc.) sont des catégories soutenues par des mécanismes homéostatiques théoriquement pertinents, soit, dans un contexte phylogénétique, les mécanismes évolutifs implicites (et explicite dans le cas de la formation de lignées) dans la définition de l'homologie. Aucune information supplémentaire à celle fournie par la définition de l'homologie n'est ajoutée par l'utilisation de la notion d'espèce naturelle.

Cela est d'autant plus important que l'injonction de l'étiquette « espèce naturelle » à considérer l'unité d'une catégorie pourtant hétérogène occulte les modifications significatives que l'homologie a subies au travers du 20^e siècle. Si la *définition* de l'homologie, aspect du concept sur lequel les théoriciens et théoriciennes des espèces naturelles focalisent leur attention, est restée plus ou moins stable, les réponses offertes au travers du temps à la question centrale « Comment savoir si deux occurrences relèvent d'un 'même' trait? » témoignent d'un dynamisme inhérent à la phylogénétique. Ces modifications, négligées par la tradition des espèces naturelles, relèvent directement de l'opérationnalisation du concept et donc des modalités de la production de connaissances phylogénétiques portant sur les molécules, les organismes et les unités taxonomiques.

5.5 Les homologues moléculaire et morphologique analysées à l'aune du cohérentisme historique

L'analyse de la similarité morphologique menant à l'inférence phylogénétique ainsi que l'analyse de la similarité entre séquences de nucléotides (des gènes si le gène est défini comme tel; Dupré 2012) structurent deux itérations du système épistémique qu'est la phylogénétique. Il s'agit en ce sens de deux procédures mobilisées par les chercheurs et chercheuses afin de répondre au problème de la coordination : comment fournir un contenu empirique à la théorie de l'évolution par voie de sélection naturelle alors que les prédictions de cette théorie concernent notamment des relations entre organismes qui sont inobservables de façon directe? Pour ce faire, il faut accepter de fait une relation entre similarité et phylogénie, mais il y a (au moins) deux façons de comprendre cette relation, l'une morphologique et l'autre moléculaire. Si les deux diffèrent sur certains points, elles partagent aussi nombre de croyances structurantes, étant toutes deux incrustées dans le cadre général offert par la théorie de l'évolution par voie de sélection naturelle. Dans ce qui suit, j'utilise les outils issus du cohérentisme historique (itérations épistémiques, impératif de progrès, principe de respect, principes ontologiques, perspective synoptique et analyse du tout centrée sur les modèles) pour analyser ces différences et similarités, ce qui m'amène à établir un constat important relatif au rôle inférentiel de l'homologie au sein des pratiques phylogénétiques, un constat dans l'angle mort de l'approche GHP.

La similarité entre la phylogénétique moléculaire et la phylogénétique morphologique se traduit par un généreux ensemble de croyances communes aux deux itérations : validité de la corrélation liant similarité et phylogénie; définitions d'unités taxonomiques opérationnelles (comme l'espèce) basées sur la notion de lignée ou d'ascendance commune; validité descriptive de la réduction d'une unité taxonomique opérationnelle à l'un de ses représentants (une séquence ou un organisme représente l'unité taxonomique dans son ensemble); validité du paradigme néo-darwinien; validité du principe de divergence et préséance des événements de spéciation dichotomiques; étude de lignées (d'organismes ou de gènes) indépendantes qui permet de retracer de façon pertinente l'histoire évolutive du monde vivant; histoire évolutive

considérée comme étant unique, de sorte que les différents niveaux d'analyse devraient converger sur les mêmes résultats; non-pertinence de l'aspect fonctionnel des traits étudiés (dans la grande majorité des cas); non-pertinence de l'environnement des lignées étudiées; etc. Tout ce bagage commun traduit des contraintes théoriques imposées aux procédures phylogénétiques. Or, malgré l'importance de ces traits partagés, les deux itérations n'en demeurent pas moins bien distinctes, une distinction plus facilement saisissable en traitant des modèles utilisés pour l'analyse de la similarité.

Les modèles d'analyse de similarité, au niveau morphologique, n'incluent aucune hypothèse concernant les séquences de transformation des traits étudiés. Ce constat a mené Patterson (1982) à distinguer les homologies taxique (synonyme de synapomorphie et qui traduit l'opérationnalisation phylogénétique du concept) et transformationnelle (qui intègre des hypothèses au sujet de la séquence de transformation des traits étudiés). Selon lui, la cladistique doit préconiser l'étude des premières et donc définir l'homologie en conséquence, c'est-à-dire sans référer aux séquences des transformations subies par les caractères. Cette définition a cependant été considérée comme étant problématique puisqu'elle semble enlever toute teneur évolutionniste à l'homologie (Brigandt 2003; Brower 2015; Carine et Scotland 1999; Griffiths 2007; Wagner 1994).

Considérons la critique de Griffiths. Selon lui, la notion taxique de l'homologie est incomplète puisqu'elle n'inclut aucune explication théorique de la similarité liant deux homologues :

On the taxic conception, my arm is homologous to the wing of a bird because birds and mammals have a common ancestor that had forelimbs. Both arm and wing are copies of copies of copies of the same ancestral limb. Homology is thus defined recursively in terms of a relation of 'copying' between the parts of parents and the parts of offspring, but no account is given of this copying relation. (Griffiths 2007, 649)

De ce point, il conclut que la définition dite biologique de l'homologie par Wagner (1989, 1994) est plus adéquate dans un contexte évolutif. Wagner insiste sur la présence de mécanismes développementaux qui explique la récursivité des homologues. Les homologues sont hérités d'un ancêtre commun et leur récursivité est le fait de la transmission de ces mécanismes développementaux : « In addition to plugging the hole in the taxic account of homology, the sort

of developmental account of homology Wagner envisages would provide a mechanistic explanation of the success of the traditional criteria of homology in identifying stable phenomena. » (Griffiths 2007, 649).

Cette réflexion entourant l'homologie est très intéressante, mais Griffiths mésinterprète les raisons derrière la priorisation de l'homologie taxique dans le cadre de la cladistique. En effet, Patterson distingue l'homologie taxique de l'homologie transformationnelle afin de cerner les modalités de l'opérationnalisation de la notion d'homologie dans le cadre de la phylogénétique (son but étant de circonscrire l'homologie pour qu'elle puisse être associée exclusivement à des groupes monophylétiques). Selon lui, l'explication des mécanismes développementaux et des autres processus évolutifs est un objectif louable, mais ce n'est tout simplement *pas* celui de la cladistique qui vise à identifier des groupes monophylétiques sur la base de la corrélation liant morphologie et phylogénie. Pour défendre ce point, Patterson détaille un exemple par lequel il démontre que différentes hypothèses relatives à la transformation de traits homologues peuvent être formulées au sujet d'un groupe d'homologues, mais que ces différentes hypothèses n'ont aucun impact sur la classification hiérarchique qui résulte de l'analyse. Il en conclut la chose suivante :

That example implies that the transformational approach to homology may be more informative, and a lot more interesting, than the taxic approach, a point I will not deny. Yet as Eldredge (1979a) argues with regard to evolutionary theory, concentrating on transformation at the expense of taxa is not fruitful. (Patterson 1982, 36)

C'est donc dire que l'analyse de la similarité, qui permet de déterminer si deux traits distincts sont des états du même caractère (des homologies), se fait sans référence à des séquences de transformation. Les modèles d'analyses de la similarité morphologiques utilisent des critères non évolutionnistes largement reconnus (par l'entremise d'une croyance bien ancrée dans la discipline) comme étant représentatifs d'une ascendance commune : critères topologique, structurel et ontogénique ainsi que présence de formes intermédiaires. Ce point est obscurci par une analyse de l'homologie qui priorise les aspects théoriques au détriment de leur opérationnalisation (on accuse alors l'homologie taxique d'être non évolutionniste en faisant abstraction du fait que les critères de similarité sont ancrés dans un schème de pensée incontestablement darwinien).

Cette discussion nous mène à constater une différence importante entre la phylogénétique morphologique et la phylogénétique moléculaire, malgré le fait que les deux soient circonscrites par un ensemble de croyances théoriques similaires. L'analyse de similarité permettant la formulation d'homologies moléculaires (je me concentre ici sur les gènes définis comme séquences discrètes de nucléotides) doit *nécessairement* inclure des modèles de transformation des séquences afin de quantifier la similarité (et la différence) entre les séquences étudiées, contrairement à l'analyse de la similarité morphologique basée sur les trois critères de Remane (auxquels on peut ajouter le critère ontogénique). Je m'explique.

Considérons l'alignement multiple de séquences nucléotidiques d'ADN formant un ensemble à l'étude²⁹. L'objectif de tels alignements est de déterminer si des séquences sont homologues, puis lesquelles, parmi un ensemble d'homologues, sont les plus similaires (et donc probablement les plus proches phylogénétiquement). Considérons trois séquences de longueurs différentes : AACTGC (séquence 1), AACTT (séquence 2) et AAGC (séquence 3). Les modèles utilisés pour l'alignement des séquences considèrent au moins trois possibilités : 1) que les nucléotides présents à un endroit soient restés les mêmes au travers du temps (A reste A, à un locus donné et dans toutes les séquences, par exemple); 2) que les nucléotides aient été substitués (C devient G, par exemple); 3) que la longueur d'une séquence ait été modifiée au travers du temps (ce qui arrive par exemple si un nucléotide présent au sein d'une séquence ancestrale est retiré des descendantes ou, au contraire, si un nucléotide est ajouté à la séquence ancestrale). L'enjeu de l'alignement est donc de savoir quels nucléotides des séquences génétiques correspondent au *même* caractère, les caractères étant des sites détenant une place

²⁹ Une réflexion similaire pourrait être effectuée par rapport à l'identification des séquences homologues qui sont par la suite étudiées par alignements multiples (cette succession d'inférences rappelle celle qui mène à la formulation des homologies premières [ou topographiques], puis à l'établissement d'homologies secondaires [ou phylogénétiques] dans le cadre de la cladistique, comme décrit au premier chapitre de cette thèse). Je me concentre ici sur les analyses des similarités impliquées dans l'alignement multiple, notamment pour alléger le propos. Je me permets ce choix par rapport au phénomène étudié (il s'agit bien de deux phénomènes épistémiques distincts) puisque la conclusion, dans le cadre de la présente discussion, est la même pour les deux : pour évaluer la similarité moléculaire de façon précise, les modèles d'analyse requièrent des hypothèses d'arrière-plan relatives aux transformations subies par les séquences génétiques.

spécifique dans les séquences et les états de caractères étant les nucléotides présents sur ce site pour une séquence donnée. La Figure 5.1 illustre la chose.

Figure 5.1 – Un exemple d’alignement de séquences génétiques

Séquence 1	A	A	C	T	G	C
Séquence 2	A	A	C	T	T	-
Séquence 3	A	A	-	G	C	-

Cette figure illustre un alignement de trois séquences. Il s’agit d’un alignement sous-optimal, mais il permet d’illustrer le fonctionnement de la chose. Chaque colonne représente un caractère et chaque rangée représente une séquence génétique. Les cases où se croisent un caractère et une séquence donnent l’état de caractère existant pour cette séquence; la similarité entre les séquences peut être évaluée sur la base de la correspondance des états de caractère pour chaque caractère étudié. Les tirets indiquent des gaps, des espaces laissés vacants dans l’alignement qui suggèrent qu’un événement d’évolution a mené à la disparition d’un nucléotide dans la séquence concernée (ou à l’ajout de nucléotides dans les autres séquences).

Le tout se complique lorsque l’on considère les différentes possibilités d’alignement. Lorsque les séquences sont de longueurs différentes, l’alignement peut en effet se faire de plusieurs façons, selon l’endroit où l’on postule que des nucléotides ont été ajoutés, retirés et substitués. Ainsi, les programmes consacrés à l’alignement de séquences (qu’ils visent à réaliser des alignements multiples, comme les programmes *Clustal* [Higgins et Sharp 1988] et *MUSCLE* [Edgar 2004], ou à dénicher de potentiels homologues d’une séquence spécifique au sein d’une banque de données, comme *BLAST* [Altschul et al. 1990]) visent à optimiser l’alignement par l’introduction d’écarts (*gaps*) au sein de certaines séquences. Or, de telles optimisations requièrent un modèle afin de quantifier la similarité et d’identifier les meilleurs alignements, ceux à préconiser pour inférer les relations phylogénétiques. Un modèle peut stipuler, par exemple, que tous les événements (insertion, délétion ou substitution d’une base) ont les mêmes chances de se produire. En ce sens, toutes les différences entre séquences auraient la même valeur (on pourrait simplement compter les nucléotides correspondants). Au contraire, un autre modèle pourrait stipuler que certains événements sont plus fréquents que d’autres (par exemple, les substitutions d’une purine [A ou G] par une autre purine ou d’une pyrimidine [C ou T] par une autre pyrimidine sont plus fréquentes que les substitutions d’une purine par une pyrimidine [ou vice versa]), les différences entre séquences qui sont dues à de tels événements devraient donc

moins compter dans le calcul de la dissimilarité que les événements plus rares. Autrement dit, certaines différences entre deux séquences, parce qu'elles sont dues à des événements plus probables, suggèrent une plus petite distance évolutive que les différences causées par des événements moins probables (selon un modèle donné de transformation des séquences). La Figure 5.2 illustre la chose.

Figure 5.2 – Deux alignements possibles pour les séquences 1 et 2

	Premier alignement						Second alignement						
Séquence 1	A	A	C	T	G	C	Séquence 1	A	A	C	T	G	C
Séquence 2	A	A	C	T	-	T	Séquence 2	A	A	C	T	T	-

Cette figure illustre deux alignements possibles pour les séquences 1 et 2. La différence entre les alignements dépend de l'endroit où sont intégrés les gaps. Afin de trouver l'alignement optimal (celui qui maximise la similarité entre les deux séquences), il est nécessaire de se doter d'un modèle qui pondère les différences entre les séquences. Un modèle qui considère que toute transformation témoigne de la même distance évolutive considérerait les deux alignements présentés ci-dessus comme étant équivalents (dans les deux cas, une mutation ou une délétion permettent de passer d'une séquence à l'autre). Au contraire, si les transformations de purines (A et G) en pyrimidines (T ou C) sont considérées comme étant plus rares (et donc témoignant d'une plus grande distance évolutive entre deux séquences), alors le second alignement est à préconiser (le même nombre de transformations est nécessaire pour passer d'une séquence à l'autre, mais l'une des transformations étant plus rare, elle témoigne probablement d'une plus grande distance évolutive entre les deux séquences et doit être évitée, si possible). Les couleurs sont utilisées simplement pour faciliter l'observation des correspondances entre états de caractères (les alignements sont habituellement illustrés de cette façon).

C'est donc dire que des modèles utilisés pour quantifier la similarité et générer des alignements doivent nécessairement mobiliser des hypothèses au sujet des modalités du changement s'opérant au sein des séquences nucléotidiques. Le fait que de nombreux modèles considèrent ces modifications génétiques comme des phénomènes largement stochastiques est le fruit d'une conjoncture sociohistorique particulière qui a fait de l'alignement un défi relevant de l'informatique plutôt que de la biologie (Morrison, Morgan et Kelchner 2015). Cette réalité ne mine cependant en rien le potentiel de remplacer de tels modèles stochastiques par des modèles basés exclusivement sur la connaissance plus fine des mécanismes de modification génétique. Cela étant dit, les modèles utilisés sont, de toute façon, imbibés de considérations sur les transformations subies par les homologues.

Ainsi, l'analyse épistémologique des méthodes d'analyse de la similarité qui permettent la formulation d'homologies, d'abord dans une perspective morphologique, puis dans une perspective moléculaire, offre un regard détaillé sur l'inclusion de la notion d'homologie au sein de la structure inférentielle de la phylogénétique. Au sein des deux itérations épistémiques considérées (morphologique et moléculaire), le rôle théorique de l'homologie est le même, soit de lier similarité et phylogénie au sein d'un schème de pensée arborescent et évolutionniste. Dans la pratique, les procédures d'observation de l'homologie sont drastiquement différentes. Le type de données mobilisées y est pour beaucoup, mais j'ai cherché à identifier une autre différence majeure au sein des modèles permettant l'observation de l'homologie. Au niveau morphologique, les modèles n'impliquent aucune information relative aux transformations subies par les traits. Il ne s'agit que de liens de similarité qui portent potentiellement un signal phylogénétique, ce potentiel ayant été établi par les succès des itérations passées et non par une explication théorique des transformations subies. Au niveau moléculaire, au contraire, les modèles utilisés pour quantifier la similarité entre séquences sont paramétrés sur la base des connaissances disponibles quant aux processus de transformation des séquences génétiques. Voilà comment deux itérations distinctes manipulent des faits « observables » pour étudier des faits inobservables. Autrement dit, voilà comment deux itérations phylogénétiques négocient le rapport entre théorie et empirie malgré le problème de la coordination.

5.6 Conclusion

Dans ce chapitre, j'ai présenté la transition vers les données moléculaires de la deuxième moitié du 20^e siècle qui a grandement contribué à déterminer les pratiques phylogénétiques telles qu'on les connaît aujourd'hui. Cette transition permet d'illustrer le modèle hélicoïdal du cohérentisme historique : la phylogénétique morphologique et celle axée sur les données moléculaires sont des itérations épistémiques dont les croyances et pratiques constitutives peuvent être comparées. D'une telle comparaison, j'ai identifié, au milieu de structures inférentielles largement partagées par les deux itérations, une différence importante au niveau de l'opérationnalisation du concept d'homologie. La phylogénétique moléculaire mobilise des modèles pour l'analyse de la similarité génétique qui incluent des croyances relatives à la

transformation des séquences, alors que les critères de similarité morphologique n'impliquent aucune hypothèse relative à la transformation des traits comparés.

En contraste avec l'analyse que j'ai effectuée, la théorie GHP des espèces naturelles (celle qui est la plus souvent utilisée pour traiter de l'homologie) focalise son propos sur les définitions de l'homologie et leur arrimage avec le cadre théorique darwinien. Cette analyse, qui utilise l'arche des espèces naturelles comme modèle épistémologique de base, adopte ainsi une perspective surplombante pour appréhender l'homologie (van Fraassen 2008; voir chapitre 2 de cette thèse). Elle tient pour acquis le lien similarité-phylogénie et les modalités de son opérationnalisation, de sorte que la différence marquée entre les homologies moléculaires et morphologiques passe inaperçue. Ce chapitre, en plus de présenter la transition vers les données moléculaires en phylogénétique, m'a donc permis de constater un développement significatif du concept d'homologie et la puissance des analyses épistémologiques encadrées par le cohérentisme historique. Dans les deux prochains chapitres, j'utilise ce cadre théorique pour analyser les pratiques phylogénétiques contemporaines.

Chapitre 6 – Analyse épistémologique de la pensée arborescente et de la phylogénétique traditionnelle

I would suggest that you remember she is old because
some of her enemies are dead.

- Carolyne Janice Cherryh, dans *Inheritor*

Watch out for a man whose enemies keep disappearing.

- Carolyne Janice Cherryh, dans *Deceiver*

6.1 Introduction

Comme démontré aux chapitres 1 et 5, le développement historique de la phylogénétique est adéquatement décrit par le modèle hélicoïdal issu du cohérentisme historique. Ainsi, ce champ disciplinaire peut être considéré comme une succession d'itérations épistémiques au sein de laquelle des itérations données répondent à des problèmes posés par les itérations précédentes et permettent ainsi un « progrès », une amélioration des pratiques au regard des standards du champ disciplinaire en question. Au sein de la pratique contemporaine, deux itérations se chevauchent : la phylogénétique traditionnelle, structurée autour de la pensée arborescente, et la phylogénétique centrée sur les modèles réseaux. L'objectif de ce chapitre est d'analyser la pensée arborescente et son inscription au sein de la phylogénétique traditionnelle. La pensée arborescente (*tree thinking*) réfère au cadre théorique abstrait qui structure ces pratiques scientifiques et auquel ces mêmes pratiques offrent un contenu empirique.

Afin d'explorer la pensée arborescente de façon efficace, je l'analyse à deux niveaux de généralité distincts. Dans la première partie, j'analyse la phylogénétique traditionnelle par l'entremise de textes s'inscrivant à un haut niveau d'abstraction (Delsuc, Brinkmann et Philippe 2005; O'Hara 1997). Le texte d'O'Hara (1997) présente explicitement la pensée arborescente et son rôle central pour la systématique phylogénétique. Il s'agit donc du point de départ de mon analyse. Le texte de Delsuc, Brinkmann et Philippe (2005) permet quant à lui de présenter la phylogénomique, soit la pratique phylogénétique qui mobilise les outils et données de la génomique. Ce deuxième texte, un peu plus concret que le premier, est un exemple de l'intégration implicite du bagage associé à la pensée arborescente au sein d'un programme de

recherche. Après avoir analysé en détail ces deux textes, je réalise une analyse critique de la pensée arborescente et de la phylogénétique traditionnelle en mobilisant les outils du cohérentisme historique. Cela me permet d'identifier trois amalgames d'idées au cœur de la phylogénétique traditionnelle ainsi qu'un principe ontologique (le principe d'histoire évolutive unique) et de mettre à jour la teneur fondationnaliste des discours issus de ce champ d'études.

Dans la deuxième partie, je me penche sur des itérations concrètes et récentes qui incarnent la pensée arborescente. J'analyse un débat entre deux groupes de chercheurs et de chercheuses au sujet d'un phylum d'archées récemment mis à jour par l'échantillonnage de données génétiques directement dans l'environnement. Plus spécifiquement, le débat concerne le placement phylogénétique de ce phylum. Deux options sont considérées : la première suggère que *Methanonatronarchaeia* est un groupe frère d'*Halobacteria* (Sorokin et al. 2018), alors que la deuxième place ce phylum beaucoup plus profondément dans l'arbre des archées (Aouad et al. 2019). L'analyse de ce débat démontre que les conclusions par rapport à la pensée populationnelle et la phylogénétique traditionnelle tirées dans la première partie s'appliquent bel et bien aux pratiques concrètes et récentes de ce complexe épistémique.

6.2 La phylogénétique traditionnelle : une itération épistémique de la systématique

La pensée arborescente est un cadre de référence qui a mené la systématique à devenir ce que j'appelle, tout au long de cette thèse, la phylogénétique traditionnelle. La pensée arborescente représente la dimension symbolique de la phylogénétique traditionnelle et a été explicitement théorisée par O'Hara (1988, 1997). Comme le souligne O'Hara, cette pensée s'exprime implicitement dans la grande majorité des pratiques phylogénétiques de la deuxième moitié du 20^e siècle et du 21^e siècle. Les différents travaux de phylogénétique appréhendés aux chapitres 1 et 5, portant sur l'homologie et autres aspects du processus inférentiel de la phylogénétique, incarnent cette pensée. Dans cette section-ci, j'analyse en détail deux textes qui permettent d'extraire les points saillants de cette pensée afin d'en brosser un portrait général. Je me penche d'abord sur la description qu'O'Hara (1988, 1997) donne de la pensée arborescente

avant d'explorer comment celle-ci peut s'actualiser dans un programme de recherche spécifique, soit celui de la génomique (Delsuc, Brinkmann et Philippe 2005).

6.2.1 La pensée arborescente : le cadre théorique de la phylogénétique traditionnelle

Selon O'Hara, l'histoire de la systématique au 20^e siècle se divise en deux périodes qui peuvent être conceptualisées comme des itérations de ce champ d'études : jusque dans les années 1960, la discipline était dominée par ce qu'il appelle, à la suite de Mayr, la pensée populationnelle; à partir des années 1960, la pensée arborescente (*tree thinking*) conquiert le champ disciplinaire, la systématique phylogénétique accaparant progressivement à partir de ce moment une dominance notoire au sein des pratiques de classification en biologie (Hull 1970, 2001). O'Hara suggère de saisir la teneur de la pensée arborescente par l'entremise de comparaisons, d'abord en la comparant à la pensée populationnelle qui l'a précédée, puis en la comparant à des cadres conceptuels considérés comme étant problématiques, soit les pensées dites du groupe et développementale.

Pour bien saisir le propos d'O'Hara, il faut d'abord décrire la pensée populationnelle³⁰. En suivant l'impulsion de Mayr (1959), qui a été le premier à théoriser la chose, O'Hara suggère de comprendre la pensée populationnelle en la contrastant elle aussi à un autre cadre de référence, soit la pensée typologique ou essentialiste. Selon Mayr, les naturalistes prédarwinien et prédarwinienne appréhendent le monde biologique dans une optique aristotélicienne (tributaire du modèle de l'état naturel [*natural state model*]; Sober 1980). L'essentialisme aristotélicien stipule que chaque espèce biologique peut être associée à une forme, soit un archétype

³⁰ Je décris ici la pensée populationnelle de façon à faire écho à la description qu'O'Hara en donne. Il convient cependant de noter que depuis la publication du texte d'O'Hara, la pensée populationnelle a fait l'objet d'un intérêt renouvelé, tant chez les philosophes que chez les historiens et historiennes de la biologie. Ces publications traitent toutefois de sujets connexes à celui qui est au cœur de ce chapitre (la place du modèle arborescent au sein de la phylogénétique), des sujets comme la nature de la causalité au sein de phénomènes d'évolution (Ariew et Lewontin 2004; Matthen et Ariew 2009; Millstein 2006; Rosenberg et Bouchard 2005) ou de l'exactitude du constat selon lequel les naturalistes prédarwinien et prédarwinienne étaient des essentialistes convaincu-e-s (Rieppel 2010b; Winsor 2006; Witteveen 2016, 2020). Je me permets donc de les escamoter de sorte à garder le propos du chapitre centré sur le modèle arborescent.

immatériel dont les caractéristiques forment l'essence de l'espèce. Dans ce cadre de pensée, les différences entre les individus d'une même espèce sont considérées comme des *accidents*, des dérogations de l'état qui se réalise naturellement lorsqu'aucune force ne contrevient à son expression. Alors que la physique, mobilisant elle aussi le modèle de l'état naturel aristotélicien, s'est épanouie, notamment chez Newton, par l'étude des forces agissant sur les corps de sorte à tronquer l'état naturel, les naturalistes prédarwinien et prédarwinienne (et dans une moindre mesure, ceux et celles ayant travaillé jusqu'au tournant du 20^e siècle) ont centré leurs efforts sur l'étude des formes idéalisées, des représentants types de chaque espèce (des exceptions notables sont les études des monstruosité par Geoffroy Saint-Hilaire ou de la dégénérescence des espèces par Buffon; Tort 1989). En ce sens, l'étiquette *typologique* décrit une tendance à délaissier l'analyse des différences individuelles au profit des caractéristiques de formes plus générales.

La pensée populationnelle fait tout l'inverse : il s'agit de délaissier l'analyse des formes idéales au profit des différences qui caractérisent les individus. L'ancrage historique de cette manière d'appréhender le monde biologique est habituellement associé aux travaux de Darwin (Mayr 1959; 1994), si bien que Mayr considère même qu'il s'agit d'une des innovations principales de l'auteur de *L'Origine*. Il va sans dire que les différences entre les organismes sont au cœur de la théorie de l'évolution par voie de sélection naturelle : pour qu'une telle évolution se produise, il *faut* que les individus aient des traits qui les distinguent des autres et que ces traits engendrent une différence de fitness entre les organismes (Gayon 1992; Lewontin 1970). De surcroît, si cette évolution doit effectivement mener à la divergence des lignées à partir d'une population ancestrale, l'idée d'un état naturel, d'une forme idéalisée transmise par les mâles d'une espèce à leur descendance (Sober 1980), doit être remplacée par la possibilité de voir les particularités individuelles s'accumuler d'une génération à l'autre. En conséquence, pour comprendre la sélection naturelle et son effet sur l'histoire naturelle, il faut reconnaître aux différences interindividuelles un rôle explicatif et non les délaissier sous prétexte que les traits individuels qui divergent du type seraient des accidents.

La pensée populationnelle est donc requise pour conceptualiser l'évolution par voie de sélection naturelle. Nonobstant cette importance, le cadre de pensée en question n'a atteint le

statut hégémonique au sein de la systématique qu'au début du 20^e siècle (O'Hara 1997)³¹. Parmi les avenues empruntées par les praticiens et praticiennes de l'époque, O'Hara souligne la génétique des populations, l'étude de la variation biogéographique, l'étude des processus de spéciations et d'autres événements de microévolution, des disciplines qui partagent un intérêt pour les dynamiques des populations. En bref, la généralisation de la pensée populationnelle au sein de la communauté a produit les fils qui allaient être utilisés pour tisser la toile de la Synthèse moderne de l'évolution.

La pensée arborescente, si elle est bien à distinguer de la pensée populationnelle, est à concevoir comme une couche de sédiments supplémentaires à explorer par l'archéologie du savoir et non comme un cadre qui serait venu remplacer la pensée populationnelle :

Tree thinking is in no sense a successor to population thinking, which is just as important today as ever been. Tree thinking is simply the phylogenetic counterpart to population thinking, and like population thinking it has brought a more completely evolutionary perspective (de Queiroz 1988, 1992, 1997; O'Hara 1988, 1992, 1996). (O'Hara 1997, 27.)

Cette citation souligne aussi que la pensée arborescente est intimement liée à la phylogénétique, un constat qui concorde avec le traitement que j'ai fait subir à ce champ disciplinaire jusqu'à ce point de la thèse (mais qui contraste avec la phylogénétique axée sur les réseaux telle qu'elle est étudiée au prochain chapitre). Elle permet d'appréhender des questions de macroévolution (principalement les dynamiques de divergence des lignées) que les pratiques issues de la pensée populationnelle peinent à saisir. Au nombre des accomplissements de la période hégémonique de la pensée arborescente, O'Hara évoque le développement des outils de

³¹ Pour une analyse approfondie des objections auxquelles la théorie darwinienne de l'évolution par voie de sélection naturelle a fait face dans la deuxième moitié du 19^e siècle, voir Gayon 1992, tout particulièrement le chapitre 3. La pensée populationnelle, en identifiant les différences entre les individus comme source d'explications biologiques, ouvrait la porte à l'importation d'outils statistiques en biologie. En effet, l'analyse statistique permet de savoir s'il est plausible que des différences mineures apparaissant d'abord chez seulement quelques individus puissent se généraliser au sein de la population grâce à l'avantage tout aussi mineur qu'elles procurent sur le plan du fitness. C'est en effet cet aspect de l'hypothèse qui a dû être défendu, dans une perspective statistique, face aux critiques. La mise en place de la génétique des populations, au début du 20^e siècle, a permis de clore ces débats en démontrant que la spécificité de l'hypothèse darwinienne (son accent sur les différences graduelles et minimes) est en effet réconciliable avec les principes de l'hérédité mendélienne (hérédité portant sur des traits discrets [traits phénotypiques ou génétiques]).

la cladistique, le développement de méthodes computationnelles pour la reconstruction d'arbres, le complément des données morphologique, puis leur dépassement par les données moléculaires et l'intégration d'information phylogénétique à des recherches s'inscrivant dans des domaines connexes (écologie, physiologie, embryologie, éthologie, etc.).

Afin de spécifier la teneur de la pensée arborescente, O'Hara procède encore une fois par comparaison. Ce cadre de pensée se distingue premièrement de la pensée du groupe (*group thinking*). En adhérant à la pensée du groupe, une agente ou un agent épistémique conceptualise les individus appartenant à un groupe comme des copies indépendantes les unes des autres. En systématique, cela revient par exemple à considérer toutes les espèces comme des occurrences indépendantes de la même catégorie, soit l'espèce en tant que rang taxonomique. L'objectif d'un tel angle d'approche des phénomènes (qui rappelle la tradition des espèces naturelles), que ce soit en biologie ou ailleurs, est d'identifier des points communs à toutes les occurrences de la catégorie étudiée. Cette perspective est mal adaptée à la phylogénétique, nous dit O'Hara, puisque les espèces sont, depuis Darwin, définies comme des lignées qui divergent toutes d'un ancêtre commun. Chaque espèce, plutôt que d'être une représentante indépendante de l'espèce en tant que rang taxonomique, est une branche d'un arbre au sein duquel toutes les espèces entretiennent des relations phylogénétiques les unes avec les autres.

Afin de souligner l'importance de cette conceptualisation des espèces dans le cadre de la phylogénétique, O'Hara donne l'exemple suivant. Considérons dix espèces détenant le même trait. L'identification d'un tel trait partagé n'est qu'une première étape pour la pensée arborescente et la phylogénétique, l'objectif étant d'étudier les événements évolutifs qui ont fait émerger le trait en question. Pour saisir ces événements, il faut d'abord reconstruire un arbre phylogénétique, c'est-à-dire un système qui informe au sujet des relations liant les dix espèces. Un tel arbre pourrait fournir des informations cruciales, par exemple en répondant à la question suivante : le trait étudié a-t-il émergé chez l'ancêtre commun à toutes ces espèces ou a-t-il évolué indépendamment deux fois ou plus? Ce type de questions, central à la phylogénétique, perd son sens par l'adoption de la pensée du groupe à l'égard des espèces biologiques. Pour le dire autrement qu'avec les mots choisis par O'Hara, la pensée du groupe exhorte à catégoriser des occurrences au sein de groupes, mais délaisse l'autre dimension des pratiques classificatoires,

soit l'articulation de systèmes, la mise en relation des catégories qui est le pain quotidien de la *systematique*.

Afin de compléter l'illustration de la pensée arborescente, O'Hara la compare finalement avec la pensée du développement. Cette pensée conceptualise l'évolution comme étant comparable à un processus développemental en ce sens que son aboutissement aurait été prédéterminé par les conditions initiales du monde biologique. De surcroît, cette perspective implique la supériorité de l'état final par rapport aux états préalables. En effet, tout comme l'organisme mature est considéré comme étant pleinement réalisé alors que les stades préalables sont considérés comme étant en développement, les êtres trônant au sommet de l'arbre de l'évolution sont considérés comme étant plus évolués que leurs ancêtres éteints. Dans cette optique, l'espèce humaine est souvent placée à la cime phylogénétique, implicitement considérée comme l'achèvement d'un progrès inhérent à l'évolution. En d'autres termes, la pensée développementale réitère la *scala naturae*, le principe d'organisation du monde naturel que favorisait déjà Aristote et qui s'est incarné en biologie de l'évolution au moins jusqu'à la généralisation de la pensée arborescente (Lovejoy 1936; Panchen 1992).

Au contraire, la pensée arborescente répudie l'association entre évolution et progrès développemental. Selon la croyance acceptée, tous les êtres vivants étant issus d'une communauté ancestrale unique, tout être vivant existant à un moment t fait partie d'une lignée aussi vieille que ses voisins. En ce sens, ces êtres sont tout aussi évolués les uns que les autres, tout aussi aptes à survivre dans le milieu qu'ils occupent les uns que les autres (jusqu'à preuve du contraire) et tout changement au niveau des pressions environnementales peut renverser des situations de dominance apparentes (par exemple, certaines lignées détenant actuellement une biomasse inférieure à celle des êtres humains pourraient être mieux adaptées aux changements climatiques anthropogéniques et ainsi renverser la vapeur dans les siècles à venir). C'est donc dire que l'évolution est loin d'être l'achèvement d'un développement qui tend à perfectionner les êtres vivants. Il s'agit d'un processus de mutations considérées comme étant aléatoires; ce processus est régi par l'interaction entre les organismes et leur environnement et tous les êtres y sont sur un pied d'égalité puisqu'ils sont les descendants de lignées adaptées à leur milieu.

En résumé, la pensée arborescente est une manière d'appréhender les objets d'étude de la biologie dans une perspective évolutionniste qui structure les pratiques de la systématique depuis la montée en puissance de la phylogénétique durant la deuxième moitié du 20^e siècle. Il s'agit de mettre l'accent sur les processus de macroévolution et de considérer les événements évolutifs, principalement la spéciation, qui permettent d'identifier et de représenter les relations entre les unités taxonomiques étudiées. La diversité inhérente au monde biologique est appréhendée sur la base des différences entre ces unités taxonomiques (en contraste avec la pensée populationnelle qui s'intéresse à la diversité statistique qui caractérise des groupes d'individus) et expliquée sur la base des relations phylogénétiques entre elles (en contraste à la pensée du groupe qui considère les objets d'étude [individus, espèces, rangs taxonomiques supérieurs, etc.] comme étant indépendants les uns des autres). Finalement, la pensée arborescente nous exhorte à considérer le processus évolutif sur la base de la complexité biologique inhérente aux relations entre organismes et environnement plutôt que sur la base du progrès (en contraste avec la pensée du développement qui considère l'état actuel de la biodiversité comme l'achèvement d'un processus développemental et de perfectionnement).

La pensée arborescente comporte cependant beaucoup plus d'éléments symboliques que ceux soulignés par O'Hara. Dans les pages qui suivent, j'insiste sur les différentes croyances et contraintes théoriques nécessaires à la pratique phylogénétique traditionnelle qui sont absentes de l'analyse d'O'Hara, ou du moins qui n'y apparaissent qu'implicitement sous la forme d'amalgames d'idées distinctes. Si la mobilisation de croyances implicites et amalgamées est nécessaire à la pratique de la systématique (qui ne peut se permettre de constamment remettre en perspective l'ensemble de sa toile de fond), il n'en demeure pas moins important, d'un point de vue épistémologique, de mettre à jour de tels nexus afin d'en saisir les dynamiques. Pour ce faire, la prochaine étape consiste à illustrer comment la pensée arborescente s'incarne au sein d'un programme de recherche concret. Le texte de Delsuc et ses collègues (2005) portant sur la génomique (naissante, à ce moment) et son influence sur la phylogénétique offre une fenêtre de qualité sur un tel phénomène.

6.2.2 La pensée arborescente et le tournant génomique

La génomique a exercé une influence marquée sur l'ensemble de la biologie depuis le tournant du 21^e siècle. Ce terme réfère à un ensemble de pratiques qui a pour objet d'étude le génome dans son ensemble (en contraste avec un accent mis sur des gènes isolés ou sur des ensembles de gènes restreints). La génomique a pris son essor sous l'impulsion de développements techniques ayant rendu disponibles des quantités massives de données moléculaires alors inédites (l'échantillonnage de matériel génétique microbien *in situ*, par exemple, a rendu accessibles des données au sujet d'organismes non disposés à la culture en laboratoire, voire inconnus jusqu'alors) et une capacité calculatoire toujours grandissante (l'accroissement en question est dû à des innovations matérielles, logicielles et mathématiques; Patané, Martins et Setubal 2018). La génomique a exercé une influence particulièrement marquée sur la phylogénétique, donnant naissance à la phylogénomique, un domaine de recherche à la croisée de la génomique et de la biologie de l'évolution (Delsuc, Brinkmann et Philippe 2005; Eisen et Fraser 2003; Koonin, Makarova et Wolf 2021; Patané, Martins et Setubal 2018).

Afin d'offrir un regard complet sur les pratiques contemporaines, il est important de saisir en quoi consistent la transition phylogénomique et son inscription dans ce que j'appelle la phylogénétique traditionnelle. L'article de Delsuc et ses collaborateurs (2005), publié dans *Nature Reviews Genetics*, représente une excellente porte d'entrée. En effet, les auteurs y examinent les principales méthodes de la phylogénomique, mettant ainsi en exergue les processus inférentiels de ce nouveau phylogénétique. Dans ce qui suit, je reviens sur le texte afin de souligner que la structure inférentielle de la phylogénomique, telle que ces auteurs l'interprètent, s'inscrit en parfaite continuité avec la phylogénétique traditionnelle : la similarité est utilisée pour inférer la phylogénie et la pensée arborescente est bel et bien adoptée pour encadrer ces inférences. Je profite aussi de l'occasion pour signaler les nouveautés introduites par la phylogénomique (principalement liées aux types de données utilisées)³². Le tout complète adéquatement l'analyse

³² La phylogénomique est un ensemble de pratiques diverses qui ne cadrent que partiellement dans la description que Delsuc et ses collègues en donnent (voir Charlebois, Beiko et Ragan 2003 ou Doolittle 1999 pour des regards plus nuancés sur la phylogénomique et contemporains en comparaison avec le texte de Delsuc, Brinkmann et Philippe

du texte d'O'Hara qui, du fait d'être inscrit à un haut niveau d'abstraction, contourne les enjeux épistémiques liés aux données et à leur traitement par les phylogénéticiens et phylogénéticiennes.

Selon Delsuc, Brinkmann et Philippe, les pratiques phylogénomiques répondent à un problème auquel faisaient face les pratiques phylogénétiques traditionnelles : que soient mobilisées des données morphologiques ou génétiques, les reconstructions d'arbres non concordantes se multiplient à la fin du 20^e siècle. Selon eux, cette situation découle du fait que les analyses phylogénétiques se font à partir d'ensembles limités de gènes ou de traits morphologiques, voire un seul gène ou trait. Bien qu'elles soient passées sous silence dans le texte de Delsuc et ses collaborateurs, au moins trois solutions s'offrent face à une telle problématique (les auteurs considèrent seulement la troisième option). Premièrement, il est possible de considérer que l'absence de concordance entre les reconstructions phylogénétiques signale que plusieurs histoires évolutives doivent être racontées : celles de chaque gène, de chaque trait morphologique, en plus de celles de chaque espèce (Baptiste et Dupré 2013). Les résultats non concordants seraient alors tributaires de phénomènes biologiques à explorer (par exemple, les transferts génétiques latéraux) et non des artefacts à dépasser. Delsuc et ses collègues n'évoquent pas cette avenue. Je reviens plus loin sur l'esprit moniste de la phylogénétique.

La seconde solution est celle retenue par les tenants et tenantes de la phylogénétique (ou phylogénomique) axée sur les réseaux (par exemple, Morrison 2005). Il s'agit de reconnaître que la non-concordance des arbres témoigne d'une information importante, mais qu'il est tout de même possible de réconcilier le tout au sein d'une histoire évolutive unique. Pour ce faire, il faut cependant abandonner le patron arborescent au profit des modèles réseaux (ou de modèles

2005). Cependant, mon objectif ici est de montrer comment la pensée populationnelle peut s'incarner au sein d'un programme de recherche concret plutôt que de broser un portrait exhaustif de ce vaste champ de recherche. La pensée arborescente est un cadre théorique d'actualité; le texte de Delsuc et ses collègues, doublé des textes que j'analyse dans la deuxième partie du chapitre, illustrent parfaitement la chose.

intermédiaires, comme Koonin et al. 2021 le suggèrent). Je présente cette avenue en détail au prochain chapitre avant de la comparer à la phylogénétique traditionnelle.

Finalement, la phylogénomique, telle que décrite par Delsuc, Brinkmann et Philippe, représente une troisième option disponible pour faire face à la non-congruence des arbres phylogénétiques construits à partir de gènes uniques ou d'une quantité limitée de traits morphologiques. Il s'agit de conserver les principaux éléments de la pensée arborescente intacts tout en intégrant à la démarche une plus grande quantité de données. La phylogénomique, nous disent Delsuc et ses collègues, offre deux avenues pour la reconstruction d'arbres phylogénétiques sur la base d'informations plus englobantes au sujet du génome des espèces étudiées. Dans un premier temps, les auteurs décrivent les méthodes utilisant comme matière première des séquences génétiques. Dans un second temps, ils décrivent les méthodes qui utilisent comme matière première des informations au sujet du génome dans son ensemble. Afin de saisir l'essence de ces manières de procéder, il convient de les faire contraster avec la pratique phylogénétique traditionnelle et non génomique.

6.2.2.1 Méthodes inférentielles de la phylogénétique traditionnelle

Au sein des démarches non génomiques, la première étape consiste inévitablement à recueillir des données, soit à former un ensemble de traits homologues (ici, le terme n'est *pas* synonyme de synapomorphie). Dans une perspective de phylogénétique moléculaire, cela revient à identifier un gène présent sous différentes formes au sein de différentes espèces. Les séquences candidates à l'homologie peuvent être issues d'un milieu naturel ou de banques de données publiques (par exemple, GenBank pour des séquences génétiques ou UniProt pour des séquences protéiques). Pour identifier de telles homologies, une première analyse de similarité est réalisée en alignant les séquences candidates par paires. Le logiciel le plus connu parmi ceux utilisés pour réaliser cette opération est BLAST (Altschul et al. 1990). En utilisant cet outil, d'immenses banques de données peuvent être prospectées afin d'identifier des séquences homologues.

Une fois les homologies ainsi identifiées, la seconde étape consiste à aligner l'ensemble des séquences homologues afin d'obtenir une caractérisation plus spécifique des relations de similarité qui les lient. Comme détaillé au chapitre 5, l'alignement est une opération extrêmement

complexe qui dépend entièrement des modèles d'évolution de séquences utilisés. Ces modèles déterminent notamment quelles modifications au sein d'une séquence témoignent d'une grande distance évolutive et lesquelles traduisent une distance plus limitée. Au nombre des logiciels utilisés on retrouve CLUSTAL (Higgins et Sharp 1988) et MUSCLE (Edgar 2004). Les alignements ainsi générés sont équivalents à des matrices de caractères en phylogénétique morphologique (chaque site ou colonne est un caractère). À ce point-ci, les séquences sont aussi rognées si nécessaire : certains sites, considérés peu informatifs parce que mal conservés au sein du jeu de données, sont retirés de l'alignement avant que les inférences phylogénétiques ne soient réalisées (Portik et Wiens 2021).

La troisième et dernière étape inférentielle consiste à utiliser l'alignement des séquences afin de reconstruire un arbre phylogénétique. Au moins trois grandes familles de méthodes peuvent alors être mobilisées. Historiquement, les méthodes associées à la parcimonie ont joué un rôle important bien qu'elles aient au cours du temps cédé du terrain aux approches développées plus tardivement. Ces méthodes reconstruisent les arbres phylogénétiques en évaluant le nombre de changements requis pour passer d'une séquence à l'autre. L'objectif est de minimiser le nombre total de changements requis pour le groupe d'unités taxonomiques étudié (Assis 2015; Felsenstein 2004).

Le deuxième type de méthodes consiste à utiliser une mesure de distance évolutive séparant les espèces. Concrètement, il s'agit de transformer les matrices de caractères en matrices de distances, puis d'utiliser les valeurs ainsi obtenues afin de reconstruire des arbres phylogénétiques (Felsenstein 2004). La méthode la plus connue parmi celles basées sur les mesures de distances est celle de la jonction des voisins (*neighbour joining*; Saitou et Nei 1987). Cette méthode fonctionne en identifiant la paire d'unités taxonomiques pour lesquelles l'alignement multiséquences suggère la plus petite distance évolutive (au sein de l'échantillon). Ces deux unités taxonomiques sont considérées comme étant sœurs et forment le premier embranchement de la reconstruction phylogénétique (un embranchement représenté par les deux unités taxonomiques, leur ancêtre commun et les liens qui les unissent). La distance entre

cet ancêtre commun et les autres unités de l'échantillon peut alors être calculée³³. On répète l'opération en ajoutant une troisième unité taxonomique issue de l'échantillon, soit celle dont la distance à l'ancêtre commun postulé est la plus petite, et ainsi de suite, jusqu'à ce que toutes les unités taxonomiques soient placées sur l'arbre (Felsenstein 2004; O'Halloran 2014; Saitou et Nei 1987).

Le troisième type de méthodes, très hétéroclite, regroupe les méthodes statistiques pour la reconstruction d'arbres. Ce groupe inclut notamment les méthodes liées au maximum de vraisemblance (*maximum likelihood*) et les approches bayésiennes (Felsenstein 2004). Dans les deux cas, l'arbre qui résulte du processus inférentiel est considéré comme étant le plus probable (il s'agit de la topologie qui est la plus probable, considérant l'échantillon [maximum de vraisemblance] ou de celle qui est visitée le plus fréquemment au sein d'un processus partiellement aléatoire [approches bayésiennes]).

De nombreux et dynamiques débats ont existé et existent toujours concernant ces méthodes, leur utilisation optimale et leur perfectionnement. De nombreux enjeux épistémologiques sont soulevés par la question des méthodes utilisées, mais je me permets de les contourner parce que toutes ces méthodes conservent la structure inférentielle au cœur de la phylogénétique, celle-ci étant mon objet d'étude principal : des données issues d'analyses de similarité sont mobilisées pour inférer de l'information phylogénétique, elle-même inaccessible à l'observation. Le problème de la coordination et ses conséquences épistémiques sont présents dans tous les cas. Cependant, il convient de noter que les méthodes statistiques occupent aujourd'hui un rôle dominant au sein de la pratique, favorisant ainsi le développement de solutions mathématiques à des problèmes qui sont tout de même fortement influencés par les particularités biologiques des phénomènes sous-jacents (Delsuc, Brinkmann et Philippe 2005; Morrison 2015). J'y reviens plus loin, puisqu'il s'agit d'un des points saillants de l'argument de Delsuc et ses collègues.

³³ $d(u, k) = \frac{1}{2} [d(f, k) + d(g, k) - d(f, g)]$, où u est l'ancêtre commun, k est l'unité taxonomique de l'échantillon pour laquelle on calcule la distance à l'ancêtre commun postulé, et f et g sont les membres de la première paire d'unités taxonomiques sœurs identifiées.

6.2.2.2 Méthodes inférentielles de la phylogénomique arborescente

En résumé, la phylogénétique traditionnelle pré-génomique fonctionne en trois étapes : d'abord, identifier un gène présent sous différentes formes (des homologues) au sein de différentes espèces; ensuite, générer une matrice de caractères basée sur ces homologues; finalement, reconstruire un arbre phylogénétique. Comme mentionné précédemment, deux familles de méthodes phylogénomiques sont décrites par Delsuc et ses collègues : celles basées sur l'étude de séquences et celles basées sur des données relatives au génome dans son ensemble. Ces deux familles divergent de la méthodologie traditionnelle tout juste décrite à la première étape.

Considérons d'abord les méthodes phylogénomiques basées sur l'étude des séquences. Alors que la pratique traditionnelle se limitait à l'étude d'un gène ou d'un groupe restreint de gènes, les méthodes phylogénomiques basées sur les séquences construisent des échantillons plus fournis : de nombreux gènes présents sous différentes formes au sein de plusieurs unités taxonomiques opérationnelles sont mobilisés. Cette différence caractérisant la première étape du processus a d'importantes conséquences pour la suite du processus inférentiel de la phylogénomique. Face à de tels échantillons de plusieurs gènes, deux voies s'ouvrent pour les chercheurs et les chercheuses : d'une part, il est possible de former une seule matrice de caractères (appelée super matrice) à partir des nombreux gènes identifiés préalablement et d'utiliser cette matrice pour reconstruire un arbre phylogénétique; d'autre part, il est possible de produire une matrice pour chaque gène (comme on le faisait traditionnellement), d'en inférer un arbre phylogénétique (pour chaque gène) et finalement de concaténer l'ensemble des arbres ainsi produits pour en obtenir un seul, appelé super arbre. L'objectif est le même dans les deux cas, soit de reconstruire un arbre phylogénétique sur la base d'une plus grande quantité de données génétiques, une possibilité technique qui anime l'espoir de voir les données triompher face aux résultats conflictuels qui pullulent au tournant du 21^e siècle : « Access to genomic data could potentially alleviate previous problems of phylogenetics due to sampling effects by expanding the number of characters that can be used in phylogenetic analysis from a few thousand to tens of thousands. » (Delsuc, Brinkmann et Philippe 2005, 362).

Ces méthodes phylogénomiques basées sur des séquences mobilisent les mêmes méthodes d'analyse de la similarité que les itérations précédentes de la phylogénétique; autrement dit, les mêmes programmes menant à l'alignement des séquences génétiques (BLAST, CLUSTAL, etc.) sont utilisés. De surcroît, les mêmes méthodes pour la reconstruction d'arbres phylogénétiques (maximum de vraisemblance, méthodes bayésiennes, etc.) sont mobilisées. Malgré ces similarités méthodologiques, de nouvelles procédures doivent être développées pour produire les super matrices et pour la concaténation des arbres caractéristiques de la phylogénomique (Ren, Tanaka et Yang 2009). Dans le cas des super matrices, il s'agit de produire, pour chaque espèce, une super séquence à partir de l'ensemble des gènes étudiés (Patané, Martins et Setubal 2018) ou de présenter au sein d'une seule et même matrice des données de types variés (de Queiroz et Gatesy 2007; Geisler et al. 2011). Dans le cas des super arbres, des algorithmes ont été développés spécifiquement pour les produire sur la base de plusieurs arbres préalablement reconstruits (Bininda-Emonds 2005; Sanderson, Purvis et Henze 1998). Bien qu'il s'agisse là, en toute rigueur, de nouveautés, il convient de souligner qu'il ne s'agit que d'ajustements à apporter aux méthodes déjà utilisées pour générer des alignements ou pour établir des consensus à partir d'arbres divergents (un problème déjà rencontré et surmonté en cladistique lorsque l'information de traits ne convergeait pas sur la même topologie).

Comparativement aux méthodes basées sur l'étude de séquences, la deuxième famille de méthodes phylogénomiques diverge plus significativement des pratiques phylogénétiques non génomiques. En effet, la façon d'analyser la similarité entre les unités taxonomiques opérationnelles diffère significativement. Ces méthodes, très variées, utilisent des données qui caractérisent des génomes et non des séquences. En 2005, alors que Delsuc et ses collègues publient leur article, plusieurs options sont considérées sur un pied d'égalité. Ces options utilisent le contenu génétique des génomes (la distance entre les génomes est calculée sur la base de gènes partagés, puis mobilisée pour inférer des arbres phylogénétiques), leur ordre génétique (la distance est calculée sur la base des « points de rupture », c'est-à-dire des paires de gènes qui se succèdent immédiatement dans un génome et non dans l'autre formant la paire comparée; Blanchette, Kunisawa et Sankoff 1999) et la présence de courtes chaînes de nucléotides (*DNA strings*) au sein du génome. Les deux premières approches caractérisent les génomes sur la base

de séquences génétiques préalablement identifiées comme des homologues. Elles requièrent donc des analyses de similarité génétique typiques, c'est-à-dire des analyses réalisées grâce à l'alignement de séquences. Au contraire, la troisième approche, de laquelle j'ai déjà traité au chapitre précédent, relève de la phylogénétique sans alignement (*alignment-free phylogenetics*)³⁴. Ces méthodes sans alignement représentent un exemple patent d'itération phylogénétique qui se comprend comme une négociation (plutôt féroce) entre le principe de respect et l'impératif de progrès. De surcroît, cet angle d'approche a, depuis la publication du texte de Delsuc et ses collègues, gagné en popularité et mérite donc une analyse plus détaillée portant notamment sur le rapport entre phylogénomique sans alignement et pensée arborescente.

Épistémiquement parlant, les entraves sur la voie de l'inférence phylogénétique sont nombreuses. La phylogénétique sans alignement permet de répondre à certaines de ces difficultés. L'analyse de séquences basée sur leur alignement est, par exemple, susceptible au bruit phylogénétique engendré par de hauts taux de mutation. Pour répondre à ce problème, les pratiques traditionnelles ont été concentrées sur des gènes bien conservés (et même sur des sites bien conservés au sein de ces gènes) dans un grand nombre d'espèces, le gène codant pour les sous-unités d'ARN ribosomal en étant un bon exemple (Delsuc, Brinkmann et Philippe 2005). Au contraire, les séquences et les sites ayant divergé de façon plus marquée sont typiquement exclus des échantillons, le plus grand nombre de mutations les caractérisant diluant le signal phylogénétique selon la perspective traditionnelle.

Or, cette solution contraint les phylogénéticiens et phylogénéticiennes à réduire l'histoire des espèces à l'histoire d'un gène bien conservé (ou d'une petite quantité de gènes bien conservés), négligeant du même coup une partie significative du génome des organismes concernés. La phylogénétique sans alignement, en changeant le point focal des analyses, évite ce problème. En effet, cette méthode contourne l'alignement de longues séquences pour se

³⁴ Certaines approches dites sans alignement procèdent par de micros-alignements pour comparer des génomes ou de longues séquences génétiques. La procédure est tout de même différente de l'alignement traditionnel (Zielezinski et al. 2019). Par ailleurs, je me concentre ici sur les méthodes sans alignement qui viennent le plus explicitement remettre en question des principes de la pratique traditionnelle, de sorte à faciliter l'analyse épistémologique.

concentrer sur l'identification de courtes chaînes nucléotidiques (les k -mers) qui ont tendance, du fait même d'être plus courtes que les séquences utilisées pour l'alignement, à mieux résister aux mutations affectant le génome. La phylogénétique sans alignement est ainsi capable d'extraire un signal phylogénétique à partir de données génétiques qui restent opaques face aux méthodes qui mobilisent l'alignement (Noviello et al. 2018; Vinga 2014; Zieleszinski et al. 2019), offrant du même coup une réponse à des problèmes rencontrés par des itérations passées de la phylogénétique.

Les analyses sans alignement facilitent le rapport aux phénomènes phylogénétiques d'une autre façon. L'alignement de plusieurs séquences requiert un pouvoir calculatoire qui croît rapidement avec le nombre et la taille des séquences à analyser. Les méthodes sans alignement se limitant à l'analyse de la fréquence de certains segments d'ADN et à leur localisation au sein des génomes à comparer, leur utilisation pour de grands jeux de données est plus facile (cette opération revient à mobiliser un algorithme de recherche de sous-chaîne). Cet avantage calculatoire des analyses sans alignement prend tout son sens dans le cadre d'une biologie de l'évolution confrontée à l'avalanche de données que l'on connaît (Agüero-Chapin et al. 2016; Pham, Gao et Phan 2017; Zieleszinski et al. 2019).

Ces améliorations potentielles de la phylogénomique se font toutefois au prix de certaines remises en question du canon de la phylogénétique traditionnelle. Comme mentionné au chapitre précédent, la phylogénétique sans alignement remet en question le critère d'analyse de similarité structurelle qui est mobilisé depuis l'époque prédarwinienne. Ce critère a été adapté à l'analyse de similarité génétique en stipulant que les nucléotides doivent être présents dans le même ordre (et non seulement présents) pour que leur présence témoigne d'une relation de similarité entre les séquences comparées. En étudiant des k -mers d'une taille restreinte et présents dans n'importe quel ordre au sein des génomes étudiés, la phylogénétique sans alignement remet en question ce principe, un geste épistémique ayant attiré sa part de critique. Il y a donc une négociation nécessairement impliquée dans la mobilisation de méthodes sans alignement : il s'agit de peser les avantages qu'elles procurent et la remise en question de principes éprouvés, idéalement en confrontant la méthode aux données empiriques.

6.2.2.3 Phylogénomique et stabilité de la pensée arborescente

Malgré la cassure relative à l'observation de la similarité suggérée par la phylogénétique sans alignement, je tiens à insister sur le fait que la phylogénomique, telle que décrite par Delsuc et ses collègues, s'inscrit dans la continuité de l'approche traditionnelle et pré-génomique. Les méthodes de reconstruction d'arbres phylogénétiques dont Desluc, Brinkmann et Philippe traitent sont en effet les mêmes que celles qui mobilisent une ou peu de séquences génétiques (méthodes statistiques, *neighbour joining*, etc.). À l'exception des méthodes sans alignement, toutes ces méthodes mobilisent de surcroît les mêmes outils pour l'analyse de la similarité que la phylogénétique pré-génomique. Dans le cas des méthodes sans alignement, les outils sont différents, mais le principe demeure le même : la similarité est mobilisée pour inférer la phylogénie. Il y a donc bel et bien continuité.

La dynamique ici mise en lumière correspond en ce sens au modèle hélicoïdal du cohérentisme historique. La pratique de la systématique phylogénétique est en constante évolution; il s'agit d'un complexe épistémique au sein duquel différentes itérations se succèdent, ces itérations étant toutes au moins partiellement différentes les unes des autres. La succession des itérations est influencée par différents facteurs. En effet, la phylogénétique a été influencée par des innovations matérielles, mathématiques et logicielles qui forgent l'itération qu'est la phylogénomique telle que décrite par Delsuc et ses collègues. Or, cette phylogénomique conserve de nombreux éléments de la structure générale des itérations passées. Le principe de respect doit donc être considéré comme étant important au sein de cette itération de la systématique. Comme je viens de l'étayer, ce principe cède l'avant-scène à l'impératif du progrès dans le cas de la phylogénétique sans alignement.

Ainsi, malgré l'accumulation de données, l'augmentation de la puissance calculatoire accessible aux chercheurs et aux chercheuses ainsi que l'ouverture à transformer la pratique par l'entremise de réflexions conceptuelles (par exemple, le rejet du critère structurel pour l'analyse de la similarité sans alignement), le complexe discursif étudié témoigne d'une stabilité issue de la cristallisation des principes théoriques rendus explicites, notamment par Willie Hennig. Un des faits marquants de cette stabilité est bien sûr la pérennité de la confiance dont les phylogénéticiens et phylogénéticiennes témoignent à l'égard du modèle arborescent.

Le statut inviolable que certains chercheurs et chercheuses accordent au modèle arborescent est parfaitement illustré par le texte de Delsuc et collègues. Outre le désir d'offrir un regard général sur les méthodes de la phylogénomique, leur article vise à mettre de l'avant un argument bien précis : malgré les promesses de la phylogénomique, l'accroissement du nombre de données et de la capacité à les traiter est et sera toujours incapable de remédier à certaines difficultés. Concrètement, nous disent-ils, tant que les modèles biologiques sous-jacents à la pratique entrent en conflit avec les phénomènes biologiques, des incongruités émergeront nécessairement. Puisque ces modèles doivent nécessairement mobiliser des hypothèses de simplification (*simplifying assumptions*), ils entrent bel et bien en conflit avec les phénomènes. Dans une courte section intitulée « Les sources des incohérences » (*Sources of inconsistency*), les auteurs donnent des exemples de telles simplifications couramment inscrites au cœur des modèles phylogénétiques : les substitutions de nucléotides sont indépendantes les unes des autres, les taux de substitutions sont homogènes sur l'ensemble des séquences étudiées, les taux de substitutions sont homogènes dans le temps, etc. C'est donc dire que l'injonction de Delsuc et ses collègues à sophistication des modèles de l'évolution afin de mieux rendre compte des phénomènes biologiques visent principalement les modèles portant sur la transformation des séquences.

En contrepartie, le modèle arborescent inhérent à l'ensemble des pratiques phylogénomiques décrites dans l'article n'est aucunement considéré comme une source potentielle d'incohérence. Celui-ci jouit d'une inviolabilité caractéristique de ce que Chang appelle un principe ontologique : la correspondance de ce modèle aux phénomènes évolutif est un fait indémontrable (à l'échelle du monde vivant), mais dont la remise en question est tout simplement non envisageable pour la plupart des praticiens et praticiennes (ou du moins, elle l'était peu en 2005). La phylogénomique, telle qu'elle est décrite par Delsuc et collègues, est donc bien inscrite au sein de ce qu'O'Hara appelle la pensée arborescente et incite à penser le modèle arborescent par l'entremise de la notion de principe ontologique.

Or, ce qui fait de ce modèle un point focal intéressant pour l'épistémologie, outre sa pérennité qui remonte à l'époque prédarwinienne, c'est sa remise en question progressive qui débute au tournant du 21^e siècle et qui n'a cessé de se populariser depuis. Cette situation est au

cœur du prochain chapitre. Pour l'instant, il convient de continuer l'analyse détaillée de la phylogénétique traditionnelle et de sa relation à la pensée arborescente.

6.3 Une première analyse critique : trois constats

S'il est impossible de résumer l'ensemble des pratiques phylogénétiques en raison de l'étendue du complexe épistémique en question, il est tout de même possible de cerner des textes qui mettent en exergue certains éléments clés des itérations contemporaines de la systématique biologique. Les textes d'O'Hara et de Delsuc et ses collègues, analysés dans les pages qui précèdent, en sont deux bons exemples. Selon la description qu'O'Hara donne de la phylogénétique, ses caractéristiques principales, qu'il subsume sous l'étiquette de la pensée arborescente, sont d'appréhender l'évolution à une échelle plus large que la pensée populationnelle (la phylogénétique s'intéresse aux dynamiques macroévolutives), de considérer les espèces comme des lignées en relation phylogénétique les unes avec les autres (plutôt que des occurrences indépendantes du rang d'espèce) et de répudier la téléologie inhérente à la *scala naturae* (au profit d'une pensée qui dissocie évolution et progrès). Delsuc et ses collègues, quant à eux, illustrent le tournant phylogénomique (en plus de réitérer implicitement l'adhésion au modèle arborescent), une transition qui se révèle plus quantitative que qualitative. En effet, si les données mobilisées par la phylogénomique sont certainement plus nombreuses que celles qui étaient utilisées à l'ère pré-génomique, leur traitement est réalisé par l'entremise de modèles partagés avec les itérations passées. Nous avons tout de même observé des nouveautés significatives de la phylogénomique qui témoignent d'une négociation entre le principe de respect et l'impératif du progrès (principalement sur le plan des méthodes sans alignement pour l'analyse de la similarité).

Dans ce qui suit, j'utilise les outils du cohérentisme historique afin de formuler une analyse critique des pratiques sous-tendues par les discours étudiés. Cette analyse se veut critique en ce sens qu'elle adopte la vision synoptique recommandée par van Fraassen (voir chapitre 2 de cette thèse). Plutôt que de considérer les pratiques dans une perspective surplombante sur la structure épistémique étudiée comme les praticiens et praticiennes le font *de facto*, je croise ce regard avec une perspective interne qui fait miroiter l'historicité des pratiques scientifiques. Le tout me

permet de formuler trois critiques à trois niveaux d'analyses distincts. D'abord, je mets en lumière une série d'amalgames d'idées qui forment la toile épistémique de la phylogénétique. Ensuite, je souligne l'adhésion des praticiens et des praticiennes à une forme de monisme non justifié. Finalement, je démontre que le discours de la phylogénétique témoigne d'une posture fondationnaliste concernant la construction de connaissances. Ces deux derniers points mettent en lumière un important manque à gagner : la phylogénétique, lorsqu'elle met de l'avant des réflexions conceptuelles et philosophiques, est ancrée dans une philosophie obsolète qui gagne à être dépassée.

6.3.1 Premier constat : la pensée arborescente implique une série d'amalgames

Dans cette section-ci, je mets à jour une série d'amalgames qui sont couramment véhiculés au sein du complexe discursif qu'est la phylogénétique. Lorsque nécessaire, je réfère non seulement aux textes d'O'Hara et de Delsuc et collègues analysés préalablement, mais aussi à d'autres textes qui rapportent mon analyse à un paysage intellectuel plus étendu. J'ai identifié trois amalgames³⁵ distincts. Premièrement, les idées utilisées par O'Hara pour décrire la phylogénétique de son époque sont présentées comme étant liées au modèle arborescent. Deuxièmement, la phylogénétique en tant qu'itération de la systématique est souvent associée au modèle arborescent. Troisièmement, les discours phylogénétiques tendent à associer l'étude de l'évolution (par voie de sélection naturelle ou non) à l'étude des lignées d'individus. Ces trois associations d'idées sont présentées comme si un lien nécessaire existait entre les idées concernées. Dans ce qui suit, je détaille la teneur de ces amalgames, mais j'insiste surtout sur l'indépendance des idées en question.

³⁵ J'utilise le terme « amalgame » au sens figuré pour référer à des assemblages d'éléments hétérogènes. Les amalgames dont je traite sont composés d'idées associées les unes aux autres pour différentes raisons et de différentes façons (historiquement, logiquement, implicitement, etc.). C'est la possibilité de décomposer ces assemblages d'idées et de noter l'indépendance des éléments constitutifs qui me pousse à utiliser ce terme.

6.3.1.1 Premier amalgame : modèle arborescent en opposition aux pensées du groupe et du développement

Dans le texte d'O'Hara analysé préalablement (O'Hara 1997), l'auteur utilise le terme « pensée arborescente » pour décrire un schème théorique qui guide la systématique biologique de l'époque. Selon O'Hara, le terme « arborescent » décrit un ensemble de pratiques qui appréhende les espèces en tant qu'entités historiques liées les unes aux autres et qui dissocie le progrès et l'évolution. Une telle association d'idées porte à confusion. Pour le dire autrement, le modèle arborescent est tout à fait indépendant des deux caractéristiques décrites par O'Hara.

Pour bien saisir l'indépendance mutuelle du modèle arborescent et du rejet des idées associées aux pensées du développement et du groupe, il faut se rappeler l'histoire de ce modèle (voir chapitre 3) qui précède de loin les théories de l'évolution. En effet, le modèle arborescent date de l'Antiquité et s'ancre dans le principe de division logique. Ce principe de division logique, qui chapeaute les analyses par embranchements dichotomiques, s'est cristallisé sous une forme bien précise appelée l'Arbre de Porphyre. Or, puisque ce modèle d'analyse fonctionne justement par décomposition (pardonnez-moi ce pléonasse qui rend explicite le sens du terme « analyse »), il se prête particulièrement bien à l'appréhension de classes d'objets (par décomposition d'ensemble en sous-ensembles, etc.). L'ensemble du monde vivant est l'objet d'analyse; les actes de décomposition successifs mènent aux différents rangs taxonomiques, aux espèces naturelles, et érigent ainsi des systèmes hiérarchiques (que ceux-ci représentent l'ordre naturel ou non). L'utilisation de ce principe de division logique, c'est-à-dire d'un modèle arborescent, par les naturalistes et biologistes de tout temps explique pourquoi les systèmes hiérarchiques ont occupé un rôle central en systématique (Crombie 1994; Morange 2017; Panchen 1992; Tort 1989; Verboon 2014).

À court terme, Darwin et sa révolution n'ont rien changé à cette manière d'ériger des systèmes taxonomiques (à long terme, la réinterprétation phylogénétique du modèle arborescent a mené à l'exclusion des taxons polyphylétiques; encore plus tardivement, les modèles d'évolution de séquences ou de transformation des traits sont devenus des incontournables de l'analyse de similarité à la base du projet classificatoire en biologie). Au contraire, la théorie de Darwin visait notamment à expliquer de tels systèmes. L'histoire

démontre donc que le modèle arborescent est indépendant de l'évolution par voie de sélection naturelle (au sein des travaux de Darwin, mais l'inverse n'est *pas* vrai en raison du principe de divergence, comme je le détaille plus bas).

Deux choses sautent alors aux yeux. D'abord, le modèle arborescent a été mobilisé au sein de pratiques classificatoires qui considéraient les espèces comme étant indépendantes les unes des autres. Chez Linné, par exemple, bien que la possibilité que certaines espèces soient le fruit d'hybridations soit reconnue (Wilkins 2009; Winsor 2006), l'idée générale est bien que les espèces trouvent leur origine dans l'acte de création divine (Tort 1989; Winsor 2006). Elles sont donc des occurrences indépendantes les unes des autres et non des entités historiques liées par des relations généalogiques. Les classifications de Linné ou de Cuvier, parmi d'autres, correspondent tout de même au modèle hiérarchique ancré dans le principe de division logique et donc à l'Arbre de Porphyre (sans pouvoir y être réduites), mais les divisions et embranchements issus de l'analyse représentent seulement des liens de similarité (sans valeur phylogénétique).

Ensuite, si la systématique phylogénétique d'aujourd'hui semble en effet dissocier progrès et évolution, il va sans dire que les évolutionnistes ont par le passé représenté et pensé les deux notions à l'unisson. Comme O'Hara l'explique lui-même, les naturalistes ont longtemps reconstruit des arbres évolutifs en plaçant l'être humain, associé sinon à la perfection ou du moins à la supériorité évolutive, à la cime de telles illustrations de l'évolution, l'exemple paradigmatique étant à trouver dans les travaux de Haeckel (Farris 1990; Lovejoy 1936; Panchen 1992; Stevens 1984; Tort 1983; Willmann 2003). Ainsi, O'Hara affirme avec raison que la systématique phylogénétique de son époque dissocie effectivement progrès et évolution. Il semble tout à fait légitime de dire que les pratiques phylogénétiques évitent la pensée du groupe pour parler des espèces ou d'autres unités taxonomiques opérationnelles. L'erreur logique qu'il commet est de lier ces deux idées avec le modèle arborescent par l'expression « pensée arborescente. » Cela génère un amalgame à partir d'idées qu'il est plus précis de décrire comme étant indépendantes.

Si de tels amalgames sont tout à fait acceptables au sein de la pratique phylogénétique, puisqu'ils la simplifient et la rendent plus efficace, le texte d'O'Hara est une analyse épistémologique du complexe discursif qu'est la phylogénétique de son époque. En ce sens, ces

amalgames sont problématiques pour au moins deux raisons. D'abord, de telles analyses sont problématiques parce qu'elles sont inexactes. Or, l'épistémologie se doit d'offrir des descriptions aussi détaillées que possible des discours étudiés. Ensuite, de tels amalgames peuvent miner les chances de réaliser des progrès scientifiques. Le cohérentisme historique démontre en effet que le progrès est le fruit d'une injonction à améliorer la pratique, dont la portée doit cependant être confrontée à celle du principe de respect qui stabilise la pratique. Ainsi, les pratiques tirent leur légitimité de pratiques antérieures, mais doivent inévitablement, pour faire mieux, transformer certains aspects de celles-ci. Si de telles transformations sont parfois réalisées grâce à des innovations matérielles, techniques ou mathématiques, elles peuvent aussi être le fruit de remises en question théoriques. Or, si les croyances théoriques sont amalgamées, en remettre une en question peut sembler être un geste de rébellion drastique puisque cela entraîne la remise en question l'ensemble de l'amalgame. Les amalgames peuvent ainsi être un frein aux remises en question théoriques constructives.

Les amalgames potentiellement problématiques tout juste décrits sont caractéristiques d'analyses surplombantes qui nient l'historicité des pratiques scientifiques en adoptant, sans nuances, la logique issue d'une théorie stabilisée pour étudier son propre fonctionnement. C'est exactement ce que des auteurs comme Chang et van Fraassen nous exhortent à éviter par l'entremise du cohérentisme historique. C'est malheureusement ce qui se produit avec la remise en question du modèle arborescent qui caractérise l'histoire récente de la phylogénétique : cette remise en question a été considérée par certains et certaines biologistes comme étant anti-darwinienne, voire antiévolutionniste, en raison d'amalgames non justifiés. Encore une fois, il s'agit de l'objet d'étude du prochain chapitre, mais afin de bien comprendre ce constat, il faut mettre au jour deux autres amalgames véhiculés par la phylogénétique traditionnelle.

6.3.1.2 Second amalgame : pratiques phylogénétiques et modèle arborescent

Le deuxième amalgame est celui qui unit les pratiques phylogénétiques et le modèle arborescent. Nul besoin de détailler la chose trop longuement puisqu'il existe aujourd'hui une remise en question du modèle arborescent, au sein même de la phylogénétique, qui témoigne de l'effritement progressif de cet amalgame et qui est l'objet d'étude du prochain chapitre. Il faut cependant rappeler qu'une partie significative des pratiques associe encore aujourd'hui

phylogénétique et modèle arborescent, comme s'il s'agissait d'un lien nécessaire. C'est bien sûr le cas d'O'Hara, qui considère que la pensée arborescente décrit les grands principes de la systématique phylogénétique. C'est aussi le cas de Delsuc et ses collègues qui discutent de la faillibilité de certains modèles d'évolution sans poser le modèle arborescent comme une source d'erreur potentielle. Ces deux textes trouvent des échos dans de nombreux autres, dont ceux issus du débat récent que j'explore plus loin dans ce chapitre (débat relatif au placement d'une nouvelle espèce d'archée). Les manuels de référence de la discipline réitèrent aussi cette posture : « Phylogenies, or evolutionary trees, are the basic structures necessary to think clearly about differences between species. » (Felsenstein 2004, xix). « Methods that explicitly test hypotheses of the descent of species have resulted in rigorously tested phylogenetic trees. » (Wiley et Lieberman 2011, 1).

Cet amalgame entre phylogénétique et modèle arborescent témoigne d'une acceptation non justifiée (aucune justification n'est présentée dans les textes étudiés) du principe de divergence qui sous-tend le modèle arborescent. Tout se déroule comme si les phylogénéticiens et phylogénéticiennes le considéraient comme inhérent à la théorie de l'évolution par voie de sélection naturelle. Or, l'argument darwinien, tel qu'explicité au chapitre 4 de *L'Origine*, démontre bien que la divergence est l'une des conséquences de macroévolution possibles de la sélection naturelle et non un fait en découlant nécessairement. Pour rappel, le principe de divergence est l'hypothèse selon laquelle l'action de la sélection naturelle sur des populations tend à faire diverger les lignées d'organismes, ces mêmes lignées qui deviennent par la suite des espèces, des genres, etc. La sélection naturelle est un processus d'élagage des variations mal adaptées à un environnement. Son impact sur une population peut autant entraîner la stabilité de cette population que des transformations. De surcroît, les éventuelles transformations des populations pourraient, au moins en théorie, converger plutôt que diverger. Ce qui assied, chez Darwin, la plausibilité du principe de divergence est ce que je suggère d'appeler le principe d'excentricité, soit l'hypothèse selon laquelle les individus qui diffèrent le plus de l'individu moyen d'une population tendent à avoir un meilleur fitness *en raison même* de cette différence. L'idée qui anime l'argument darwinien est la suivante : plus des individus se ressemblent, plus la compétition entre eux sera féroce puisqu'ils ont besoin des mêmes ressources, des mêmes

espaces, des mêmes partenaires, etc. Ainsi, en se distinguant des masses, certains individus et leurs descendants peuvent ouvrir de nouvelles niches écologiques au sein desquelles la compétition moins soutenue leur permet de prospérer.

Au-delà de la validité empirique de ce postulat, il est important de distinguer quatre éléments théoriques : sélection naturelle, principe d'excentricité, principe de divergence et évolution par voie de sélection naturelle. Darwin, en résumé, nous dit la chose suivante : si la sélection naturelle opère effectivement sur des populations d'organismes au sein desquelles on observe variation et hérédité et s'il est vrai que les individus qui se distinguent le plus des autres survivent et se reproduisent mieux que les autres dans un environnement donné, alors les populations doivent évoluer par voie de sélection naturelle, de sorte que les lignées d'organismes divergent de plus en plus les unes des autres. Dans un tel scénario, le modèle arborescent peut indéniablement nous aider à inférer les liens phylogénétiques et à les représenter avec précision. Mon analyse suggère aussi que l'hypothèse de l'évolution par voie de sélection naturelle est potentiellement indépendante du principe de divergence. Même s'il s'avérait que la meilleure description possible de l'histoire des espèces sur Terre mobilise le patron arborescent, il n'en demeure pas moins erroné de présenter la pensée arborescente comme le cadre à associer *de facto* à la phylogénétique. Cet amalgame occulte une adhésion non justifiée au principe de divergence qui dépasse l'acceptation de l'évolution et de la sélection naturelle.

6.3.1.3 Troisième amalgame : phylogénétique et évolution par voie de sélection naturelle

Le troisième amalgame est plus subtil, mais d'une importance capitale autant d'un point de vue épistémologique que d'un point de vue biologique. Il s'agit de traiter comme étant unies par un lien de nécessité la croyance selon laquelle l'évolution par voie de sélection naturelle est un processus qui influe l'état de la biodiversité sur Terre et celle selon laquelle il faut, en conséquence, tracer l'histoire des lignées biologiques (d'espèces, de gènes ou d'autres) afin de comprendre l'évolution. Ces deux idées semblent en effet aller de pair :

Understanding phylogenetic relationships between organisms is a prerequisite of almost any evolutionary study, as contemporary species all share a common history through their ancestry. The idea of phylogeny follows directly from the theory of

evolution presented by Charles Darwin in *The Origin of Species*: the only illustration in his famous book is the first representation of evolutionary relationships among species, in the form of a phylogenetic tree. (Delsuc, Brinkmann et Philippe 2005, 361)

Cette association d'idées résulte toutefois d'une façon bien particulière d'appréhender l'hérédité (condition nécessaire à l'évolution par voie de sélection naturelle) qui s'articule mal avec la théorie darwinienne de l'évolution telle qu'on la comprend de nos jours. Les lignes qui suivent reprennent un argument énoncé de façon plus détaillée ailleurs (Papale 2021) qui met en lumière l'arrimage maladroit entre hérédité, formation de lignées et évolution par voie de sélection naturelle, un arrimage typique de la biologie de l'évolution traditionnelle.

Depuis Aristote, l'hérédité biologique est un phénomène connu, sinon sous ce nom, du moins sous la même forme que Darwin mettait de l'avant : le semblable génère le semblable. Ce fait anodin permet à l'effet de la sélection naturelle (sélection des organismes les mieux adaptés à un milieu écologique donné) d'être transmis d'une génération à l'autre puisque les variations avantageuses ont de bonnes chances d'être transmises à la génération suivante. La logique dominante considère qu'il convient ainsi d'associer l'évolution par voie de sélection naturelle à la formation de lignées qui permettent de retracer l'histoire des traits avantageux et des organismes qui les portent.

Or, dans cette optique, l'hérédité ou la capacité de transmettre un trait à ses descendants est une capacité des individus. En ce sens, la condition nécessaire pour qu'il y ait évolution par voie de sélection naturelle est, en ce qui concerne l'hérédité, qu'une population soit formée d'entités biologiques capables de former des lignées. Cette croyance est encore bien présente de nos jours. Après tout, la phylogénétique vise à reconstruire la trajectoire de telles lignées. L'évolution par voie de sélection naturelle, plus largement, est comprise en ces termes. Dans son influent livre de 2009, Godfrey-Smith décrit par exemple l'évolution par voie de sélection naturelle comme étant réalisée au sein de populations qui témoignent notamment d'hérédité définie comme suit : « Inheritance is understood as similarity between parent and offspring, due to the causal role of the parents. » (Godfrey-Smith 2009, 39).

L'évolution par voie de sélection naturelle traduit cependant l'évolution de populations et non l'évolution d'individus. Ce point est largement accepté, notamment par Godfrey-Smith. En

effet, l'évolution est souvent comprise comme étant le changement (ou la stabilité) dans la distribution de la variation au sein d'une population. L'évolution par voie de sélection naturelle traduit quant à elle le changement de la distribution de la variation au sein d'une population qui est régi par des pressions sélectives (Bourrat 2014; Charbonneau 2014; Papale 2021). Pour qu'il y ait évolution par voie de sélection naturelle, l'hérédité est effectivement nécessaire. Sans elle, l'état de la distribution de la variation d'une génération pourrait être aléatoire par rapport à la génération précédente; les pressions sélectives, même dans un environnement stable, seraient alors incapables de diriger les transformations subies par la population (et pourraient ainsi mener à des adaptations). Or, cette description du rôle théorique assigné à l'hérédité témoigne qu'il s'agit bien d'une condition de possibilité de la sélection naturelle qui réside au niveau des populations et non à celui des individus. Pour reprendre le terme utilisé par Charbonneau (2014) et afin de minimiser l'ambiguïté sémantique, on peut appeler *mémoire* ce trait des populations dont l'état de la distribution de la variation à un moment t est non aléatoire par rapport à l'état de la distribution de la variation au moment $t - 1$ (c'est la définition que j'en donne dans Papale 2021). Tant que cette condition est réalisée³⁶ au sein d'une population, alors l'hérédité nécessaire à l'évolution par voie de sélection naturelle est présente (ce qui ne garantit pas qu'il y ait évolution par voie de sélection naturelle).

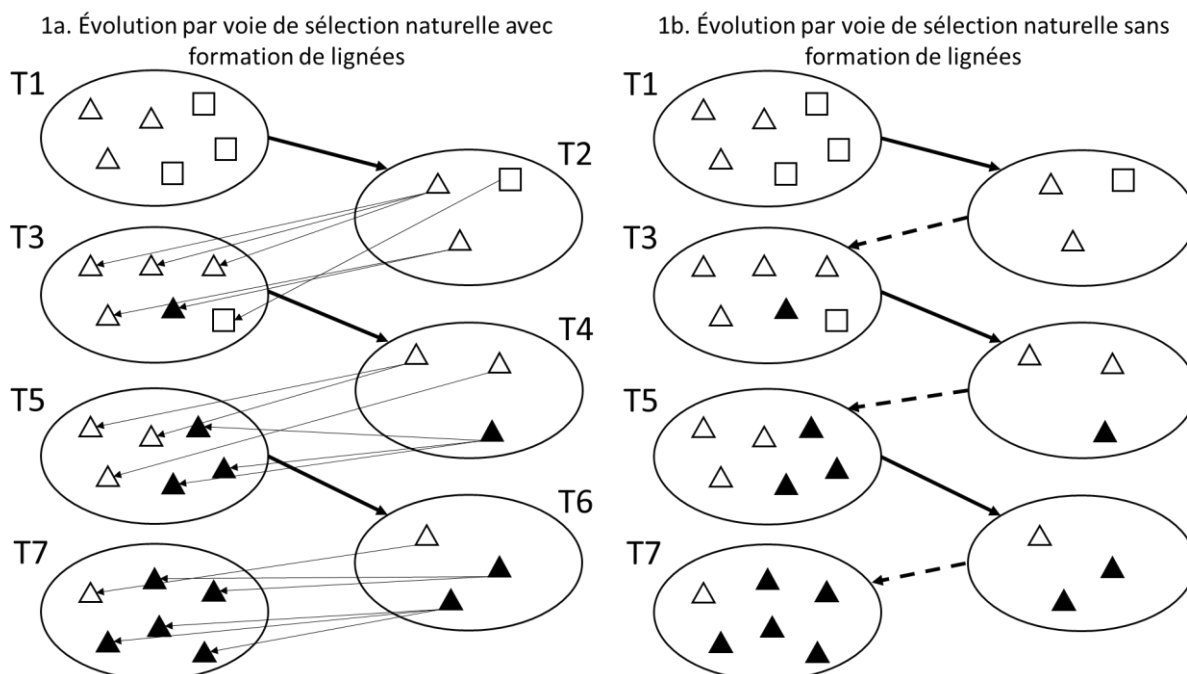
Le point important qui reste à ajouter est le suivant : la mémoire peut être réalisée de différentes façons. Si des individus produisent des rejetons qui leur ressemblent (s'ils se reproduisent), alors la population qu'ils forment devrait satisfaire la condition de la mémoire. Dans ce cas précis, des lignées sont bel et bien formées et elles sous-tendent les dynamiques darwiniennes. Si des individus sont produits (et non *reproduits*) de sorte que les effets passés de la sélection naturelle se reflètent dans les générations subséquentes et que cette production est *réalisée sans que les individus des générations passées ne jouent un rôle causal privilégié et déterminant sur cette production*, alors la condition de mémoire sera tout de même remplie. Dans ce second cas de figure, l'hérédité est une affaire populationnelle et non une affaire individuelle.

³⁶ Il serait plus adéquat de décrire l'hérédité et la mémoire par des descriptions mettant de l'avant la gradualité : une population peut présenter plus ou moins de mémoire. Or, afin de simplifier le débat, je parle ici plus simplement d'absence ou de présence du trait en question.

Il n'y a aucun lien phylogénétique entre les générations, et l'évolution par voie de sélection naturelle est tout de même possible, comme l'illustre la Figure 6.1.

Ainsi, la croyance selon laquelle la bonne compréhension de l'évolution requiert une bonne connaissance de la phylogénie des unités taxonomiques opérationnelles est issue d'une confusion entre deux idées : l'hérédité comme condition nécessaire à la sélection naturelle et l'hérédité telle qu'elle est réalisée par la formation de lignées. Il est important de les distinguer. En ce sens, il se pourrait très bien que l'étude de lignées biologiques (la phylogénétique) et l'étude de l'évolution par voie de sélection naturelle soient des tâches indépendantes l'une de l'autre. Autrement dit, les populations qui évoluent par voie de sélection naturelle pourraient être distinctes des populations constituées d'entités (gènes, organismes, etc.) qui forment des lignées divergentes.

Figure 6.1 – Évolution par voie de sélection naturelle avec et sans formation de lignées



Cette figure représente deux scénarios similaires et idéalisés d'évolution par voie de sélection naturelle. En 1.a, la mémoire (hérité au niveau populationnel) est réalisée par la formation de lignées. En 1.b, la mémoire est réalisée sans formation de lignées. Les ovales représentent des populations. Il y a une population dans chaque panneau (1.a et 1.b) qui subit des transformations évolutives au travers du temps (T1 à T7). Les flèches épaisses, allant de gauche à droite, présentent des événements de sélection qui entraînent des changements dans la distribution de la variation au sein de la population. Les flèches minces et continues (en 1.a) représentent une multiplication des individus formant la population, une multiplication réalisée de sorte que des lignées soient formées. Les flèches discontinues (en 1.b) représentent des événements de multiplication des individus au sein de la population qui n'impliquent aucune formation de lignées (une cause extérieure vient produire des individus). Dans les deux cas, les événements de multiplication permettent l'apparition de variation. Dans les deux cas, les populations font preuve de mémoire puisque la distribution de la variation, au travers du temps, est non aléatoire relativement aux états passés de la population. Dans les deux cas, l'évolution par voie de sélection naturelle peut être réalisée (si d'autres conditions, outre la mémoire, sont satisfaites; voir Papale 2021 pour le détail).

6.3.1.4 Un bref retour sur le premier constat

La pensée arborescente est composée d'une série d'idées qui entretiennent des relations complexes. Cet amalgame est utile pour simplifier la pratique, mais une analyse épistémologique préconisant une vision synoptique permet de croiser les règles logiques qui prévalent en phylogénétique avec l'historicité des croyances sous-jacentes. Cela permet notamment d'isoler des sous-ensembles de croyances amalgamées malgré leur indépendance historique et logique. Ces amalgames peuvent être nuisibles s'ils compliquent la remise en question constructive

d'éléments des itérations passées de la phylogénétique; pour l'épistémologie, ils sont nuisibles d'un point de vue descriptif.

En bref, les trois amalgames mis en lumière présentent des principes chers à la phylogénétique contemporaine comme étant liés de façon nécessaire. Premièrement, dans le texte d'O'Hara, le modèle arborescent est opposé explicitement à la pensée du groupe et à la pensée du développement (il est amalgamé avec leurs opposés) alors qu'il est compatible avec ces deux façons d'appréhender le monde biologique. Deuxièmement, le modèle arborescent est régulièrement présenté comme allant de pair avec l'évolution par voie de sélection naturelle alors que les deux peuvent se passer l'un de l'autre. Troisièmement, la pensée arborescente suggère que l'évolution par voie de sélection naturelle est un phénomène nécessairement lié à la formation de lignées alors que cette association relève plutôt de la conjoncture historique ayant lié les deux que de l'analyse logique rigoureuse.

6.3.2 Deuxième constat : le modèle arborescent est un principe ontologique

De nombreux et nombreuses biologistes, dont O'Hara et Delsuc et ses collègues, considèrent que le modèle arborescent est plus qu'un outil de recherche. La topologie arborescente traduisant l'histoire de l'évolution sur Terre serait une réalité qui peut être découverte, qui existe indépendamment de l'esprit humain et de ses activités scientifiques. Cette croyance est particulièrement importante puisqu'elle figure à la base de diverses justifications épistémiques de la discipline. En effet, la validité des méthodes utilisées, des hypothèses d'homologie et des topologies phylogénétiques est établie en mobilisant le principe de congruence, lui-même dépendant d'une posture bien précise à l'égard de l'arbre phylogénétique. Comme je le démontre dans les lignes qui suivent, cette logique témoigne du fait que cet arbre est considéré conjointement à un principe ontologique (Chang 2004). En ce sens, et c'est le deuxième constat que j'extrai de l'étude épistémologique de la pensée arborescente et de la phylogénétique traditionnelle, ce champ d'études adopte une posture fondationnaliste par rapport à la construction de ses propres connaissances, une posture qui repose sur une croyance

non justifiée. Afin de bien saisir toute la portée de ce constat, il convient de revenir sur la notion de principe ontologique.

Un principe ontologique est une croyance acceptée *a priori*, injustifiée et parfois injustifiable (Chang 2004). Ces croyances sont tout de même nécessaires aux pratiques épistémiques. Il s'agit de principes basaux qui rendent les inférences possibles si elles sont acceptées de façon apriorique comme prescrit par la culture intellectuelle concernée. Les principes ontologiques se distinguent des autres croyances acceptées de façon apriorique, notamment par la profondeur de leur ancrage : ils sont rarement explicités et difficilement remis en question.

L'exemple mobilisé par Chang pour illustrer cette notion (et du même coup, pour lui offrir un contenu empirique) est celui du critère de comparabilité. Regnault, au 19^e siècle, a démontré la supériorité du thermomètre à l'air par rapport à celui au mercure en les évaluant sur la base de ce critère. Pour rappel, la comparabilité est l'idée selon laquelle un instrument de mesure donné devrait obtenir des résultats comparables à chaque utilisation réalisée dans des conditions similaires. La comparabilité est considérée par Regnault comme une vertu des instruments testés en raison du principe de quantité unique, soit l'idée qu'un corps, dans un contexte donné, ne devrait avoir qu'une valeur pour chaque quantité théorique mesurée. Ce principe de valeur unique est un principe ontologique : pour en démontrer la validité, il faudrait avoir un accès indépendant aux valeurs à quantifier, mais un tel accès n'est disponible que par l'entremise des outils dont on cherche à établir la validité (les thermomètres, pour Regnault). Nonobstant cette indémontrabilité, le principe de valeur unique structure la pratique en contraignant le champ des possibles : une procédure métrologique qui se révèle moins régulière qu'une autre (qui ne passe pas le test de la comparabilité) tend à être rejetée.

La phylogénétique, caractérisée par la pensée arborescente, exploite un principe du même acabit, que je suggère d'appeler le principe d'histoire évolutive unique. Lorsque j'ai analysé le texte de Delsuc, Brinkmann et Philippe, j'ai démontré que ces auteurs, bien qu'ils reconnaissent la faillibilité de certains modèles phylogénétiques (tout particulièrement ceux relatifs aux transformations subies par les molécules étudiées en phylogénomique), ne mentionnent pas la

possibilité que le modèle arborescent soit source d'incongruences ou d'erreurs. Son acceptation est apriorique et bien ancrée; il est largement considéré que ce modèle correspond à *la* topologie réelle de l'histoire évolutive du monde vivant : « A basic principle of phylogenetic realism is that homologies and phylogenies require discovery. » (Wiley et Lieberman 2011, 123). Delsuc et ses collaborateurs vont jusqu'à considérer qu'une bonne compréhension des phénomènes phylogénomiques requiert la découverte de *la* topologie phylogénétique arborescente (et non d'une topologie pour chaque gène ou chaque niveau d'organisation, par exemple) : « Assembling the tree of life represents the first step towards achieving the big picture of phylogenomics where, to paraphrase Theodosius Dobzhansky, nothing in genomics makes sense except in the light of evolution. » (Delsuc, Brinkmann et Philippe 2005, 373). Cette croyance est largement partagée au sein de la phylogénétique :

Charles Darwin contended more than a century ago that all modern species diverged from a more limited set of ancestral groups, which themselves evolved from still fewer progenitors and so on back to the beginning of life. In principle, then, the relationships among all living and extinct organisms could be represented as a single genealogical tree. Most contemporary researchers agree. (Doolittle 2000, 90)

Cette adhésion au modèle arborescent doublé du principe d'histoire évolutive unique implique un ensemble de contraintes théoriques qui guident la pratique phylogénétique (comme illustré dès le premier chapitre de cette thèse). Le modèle arborescent insiste par exemple sur la préséance d'événements de spéciation dichotomiques (Hennig 1965; Panchen 1992; Rieppel 2016), éléments constitutifs principaux des topologies phylogénétiques considérées comme viables (au sein de la phylogénétique traditionnelle). Les topologies impliquant de la réticulation sont au contraire exclues *de facto* (toujours dans la perspective traditionnelle et à l'exception de certains événements rares comme l'endosymbiose ayant donné naissance aux eucaryotes).

Le principe d'histoire évolutive unique joue d'autres rôles épistémiques que celui de contraindre les topologies phylogénétiques par son alliage au modèle arborescent. Considérons par exemple le test de congruence, décrit au premier chapitre de cette thèse, qui permet d'évaluer la validité d'une hypothèse d'homologie formulée à l'égard de traits morphologiques. Si une ressemblance morphologique est effectivement le fruit d'une relation d'homologie, alors elle devrait concorder avec les autres homologies observées pour le groupe d'espèces étudié, et

donc avec l'histoire évolutive sous-jacente (Assis 2015; de Pinna 1991; Rieppel 2015). D'où vient la confiance en ce test? On peut se fier au test de congruence si et seulement si on entretient une croyance selon laquelle il n'y a qu'une seule histoire évolutive à raconter, à découvrir, et donc que tous les signaux phylogénétiques valables (comme les homologies) doivent converger sur cette histoire. Cette croyance est d'ailleurs régulièrement réitérée sous diverses formes qui dépassent le test de la congruence des homologies : « When two phylogenies are fundamentally discordant, at least one data set must be misleading. » (Losos, Hillis et Greene 2012, 1429).

La congruence permet aussi de valider de nouveaux types de données et des méthodes inférentielles. En effet, lorsque les données moléculaires sont devenues accessibles, leur légitimité a notamment été établie par la démonstration qu'elles pouvaient mener à des résultats congruents avec ceux préalablement obtenus sur la base de données morphologiques (voir chapitre 5; Inkpen et Doolittle 2016; O'Malley 2016). De la même façon, lorsque des nouvelles méthodes d'inférence phylogénétique (Felsenstein 2004; Panchen 1992) ou d'analyse de la similarité (Bernard et al. 2017; Zielezinski et al. 2017, 2019) sont développées, leur validité est établie par congruence : « The congruence of results obtained from various datasets and/or various methods is the key validation of evolutionary inferences. » (Delsuc, Brinkmann et Philippe 2005, 372). Ce rôle que la congruence joue en phylogénétique dépend non seulement de la validité d'un principe ontologique, soit celui de l'histoire évolutive unique, mais aussi du principe de respect : même si l'on cherche à dépasser les itérations passées pour « faire mieux », les nouvelles itérations tirent au moins une partie de leur légitimité des précédentes.

Ainsi, le principe d'histoire évolutive unique régit effectivement les possibilités inférentielles des praticiens et des praticiennes qui l'adoptent (par exemple, en validant le test de congruence et en décourageant le pluralisme explicatif ou descriptif). De plus, sa validité est indémontrable dans l'absolu puisque l'observation des relations phylogénétiques est soumise au problème de la coordination; seul le principe de respect (et donc les succès passés réalisés sur cette base) justifie l'adhésion à ce principe. Il y a donc bel et bien une aura particulière entourant l'Arbre phylogénétique, au singulier, qui est considéré comme étant une réalité à découvrir : « Hypotheses about relationships among organisms are meant to estimate the true phylogenetic tree that exists in nature at an appropriate level of complexity. [...] The fact that there is only one

true tree at this level of complexity provides the basis for testing alternative hypotheses. » (Wiley et Lieberman 2011, 6). Le tout suggère indéniablement l'existence d'une posture réaliste et moniste bien ancrée au sein de la discipline.

6.3.3 Troisième constat : l'ancrage philosophique de la phylogénétique est fondationnaliste et donc obsolète

La discussion au sujet de la congruence fait écho aux débats concernant le réalisme en philosophie de la mesure exposé au chapitre 2. J'y décris la posture d'Isaac (2019) qui considère notamment que la convergence de différentes méthodes métrologiques sur les mêmes points fixes justifie une perspective réaliste à l'égard de ces points fixes. Cette posture réaliste est aussi fondationnaliste en ce sens qu'Isaac considère qu'un fait au sujet de la réalité, soit l'existence de points fixes, nous permet d'expliquer un fait épistémologique, soit la convergence des méthodes. Or, cette explication épistémologique de la convergence des méthodes est invalide : elle considère les points fixes comme une cause; l'observation rigoureuse des itérations épistémologiques concernées témoigne au contraire du fait que ces points fixes sont un effet de la pratique, de sorte que postuler leur existence indépendante de la pratique (ce qui est nécessaire pour en faire des causes de la convergence) est inadéquat. La convergence des méthodes est un objectif des praticiens et des praticiennes qui forgent consciemment les méthodes de manière à l'obtenir (Chang 2004; Tal 2016).

Le rôle du principe d'histoire évolutive unique ayant maintenant été mis à jour, la discussion sur la convergence en philosophie de la mesure peut être transférée à la philosophie de la phylogénétique. L'utilisation de ce principe ontologique témoigne non seulement d'une posture réaliste et moniste largement partagée par les phylogénéticiens et phylogénéticiennes, mais aussi d'une tendance à adhérer, explicitement ou non, au fondationnalisme. Au sein de cette discipline, la validité des connaissances est expliquée philosophiquement, tout aussi naïvement que dans le conventionnalisme (une actualisation du fondationnalisme) de la première moitié du 20^e siècle, comme étant le fruit d'un rapport unidirectionnel entre empirie d'un côté et théories, modèles ou hypothèses scientifiques de l'autre. L'ensemble de cette thèse est une affirmation du

fait que la situation est beaucoup plus complexe que le fondationnalisme le suggère, mais voyons tout de même comment celui-ci s'incarne en phylogénétique.

Un excellent exemple de fondationnalisme au sein de la phylogénétique traditionnelle se trouve dans la pensée de la preuve intégrale (*total evidence*). Introduite en systématique par Kluge (1989) et directement inspirée des travaux de Carnap (1950), cette pensée s'inscrit au sein d'une conception poppérienne de la systématique (Kluge 1998; Rieppel 2005a, 2009) dans le but d'argumenter contre certaines pratiques phylogénétiques, comme la pondération différentielle des traits utilisés, le traitement indépendant de différentes parties d'un jeu de données ou le rejet apriorique de certaines observations. Kluge affirme que la mobilisation du maximum de données menant à la soi-disant « preuve intégrale » permet de réduire la quantité de connaissances d'arrière-plan nécessaires et ainsi de conformer la pratique phylogénétique à certains idéaux épistémologiques. En argumentant contre la pondération différentielle des traits, il suggère effectivement la chose suivante : « the assumptions required to justify the expected weights add to background knowledge, which decreases the probability of cladistic hypothesis in light of its tests. » (Kluge 1998, 155). Au contraire, la convergence des données non filtrées sur une seule et même topologie arborescente garantirait le meilleur rapport entre théorie et phénomènes phylogénétiques.

En bref, la logique qui sous-tend la pensée de la preuve intégrale est la suivante : une plus grande quantité d'observations minimise les croyances d'arrière-plan mobilisées, de sorte à favoriser les inférences de qualité; inversement, filtrer les données en amont requiert l'utilisation de croyances d'arrière-plan, ce qui « réduit les probabilités des hypothèses cladistes ». Ironiquement, le tout est inscrit au sein d'une philosophie de la réfutabilité d'inspiration poppérienne (Kluge traite les homoplasies comme étant des observations qui minent la validité d'hypothèses phylogénétiques, qui les réfutent). Popper était toutefois très conscient de l'inévitabilité de la charge théorique des observations. Sa philosophie de la réfutabilité suggérait qu'il était possible, pour des fins d'analyses, de mettre des croyances d'arrière-plan de côté, mais qu'il était impossible de s'en défaire complètement. Thornton nous dit de Popper la chose suivante : « he emphasises both the importance of questioning background knowledge when the need arises, and the significance of the fact that observation-statements are theory-laden »

(Thornton 2021). Les observations suggérées par Kluge, aussi nombreuses soient-elles, peuvent difficilement réduire la quantité de bagage apriorique menant à l'inférence phylogénétique, du moins dans une perspective poppérienne. En ce sens, la philosophie de la preuve intégrale s'inscrit plus adéquatement au sein du positivisme logique, et donc en cohérence avec le conventionnalisme d'auteurs comme Reichenbach, qu'au sein des travaux de Popper (Kearney et Rieppel 2006).

Bien que ce genre d'approche qui octroie aux données un rôle de fondation à la connaissance soit commun en phylogénétique, des voix s'élèvent régulièrement pour souligner que les données ne peuvent prendre sens que par l'entremise d'un travail théorique fait en amont. Par exemple, les détracteurs et détractrices de la preuve intégrale (Assis 2015; Kearney et Rieppel 2006; Mooi et Gill 2010) mettent de l'avant l'idée selon laquelle la mobilisation de grandes quantités de données n'assure aucunement la validité des résultats. Il faut plutôt choisir quelles données sont utiles pour inférer les liens phylogénétiques, ou autrement dit, quelles croyances d'arrière-plan devraient être mobilisées pour acquérir et filtrer les données qui correspondent le mieux à nos objectifs théoriques. Le critère quantitatif inscrit au sein de la preuve intégrale est, selon eux et elles, inadéquat si pris seul : il transforme un problème biologique en problème algorithmique (l'analyse des matrices de caractères) dépourvu de contenu empirique (Mooi et Gill 2010) et il exagère l'importance du test de congruence pour l'évaluation des homologies au-delà de ce qui est *théoriquement* justifié (Assis 2015; Kearney et Rieppel 2006). Dans tous les cas, leurs arguments remettent à l'avant-plan l'étude qualitative des traits utilisés pour évaluer le potentiel homologique des relations de similarité liant ces traits (une réflexion qui rappelle l'injonction de Delsuc et ses collègues [2005] à améliorer les modèles d'évolution des séquences génétiques). La citation suivante illustre la force de ces critiques :

It should come as no surprise that we urge a careful review of methods in modern molecular systematics when pumpkin pie can be shown to be the sister taxon to a crayfish with 100% bootstrap support [...]. Optimizations are not so much phylogenies as they are odes to algorithms, ways to explore statistics rather than ways to explore biology. Optimization does not differentiate between homology and nonhomology; resulting topologies are 'solutions' to datasets that have not taken the foundations of phylogenetics into account. (Mooi et Gill 2010, 38)

Il semble y avoir une réflexion épistémologique en phylogénétique qui exhorte à éviter les écueils du fondationnalisme et les héritages d'un positivisme postulant la possibilité d'observations non chargées de théorie. Les critiques de la preuve intégrale s'écartent du fondationnalisme en ce sens qu'ils et elles évitent de considérer naïvement les données (*data*) comme des justifications données (*given*) des hypothèses et théories. Ce faisant, certaines de ces critiques ouvrent la porte au rejet du traitement du modèle arborescent en tant que principe ontologique :

Some contemporary systematists interpret the meaning of total evidence as the requirement of giving the phylogenetic tree complete logical priority over critical comparative studies of character data – from such a viewpoint, it is only the tree, not empirical character evaluation, that can inform us about homology and what a legitimate character might be. (Kearney et Rieppel 2006, 370)

La priorité logique de la congruence topologique des données sur une hypothèse arborescente précise doit être rejetée au profit d'une analyse théoriquement chargée des données mobilisées. Cette intuition cadre parfaitement avec les prescriptions du cohérentisme historique : les observations étant théoriquement chargées et la théorie s'étant construite par l'entremise d'itérations successives au sein desquelles un rapport d'influence mutuelle entre théorie et empirie est omniprésent, il convient d'établir la validité inférentielle d'une structure épistémique non pas sur la base de la possibilité que les données nous offrent des réponses athéoriques, mais plutôt en se demandant « pourquoi regardons nous les phénomènes de telle ou telle façon? »

6.3.4 Un bref retour sur cette première analyse critique

Dans ce qui précède, j'ai mis de l'avant trois constats qui émanent d'une analyse critique de la phylogénétique traditionnelle et de la place qu'y occupe la pensée arborescente. Pour que ces constats prennent sens, il faut abandonner la perspective anhistorique qui accepte *de facto* la logique du cadre théorique stabilisé qu'est la pensée arborescente. La perspective synchronique offre un nouveau regard sur des associations d'idées qui semblent nécessaires en adoptant une perspective surplombante. Ces idées sont en fait indépendantes autant d'un point de vue logique (en adoptant une logique autre que celle de la pensée arborescente) que d'un point de vue historique. J'ai donc suggéré une série d'amalgames qui figurent au cœur de la

phylogénétique traditionnelle et qui gagnent à être décortiqués autant pour faciliter le progrès scientifique que pour assurer des descriptions épistémologiques précises.

J'ai aussi souligné que le rôle occupé par le modèle arborescent au sein de la phylogénétique traditionnelle est celui d'un principe ontologique, soit d'une croyance apriorique injustifiée qui structure une partie significative des inférences du champ gnoséologique et qui est rarement remise en question. La présence de ce principe ontologique témoigne du fait que la phylogénétique traditionnelle tend à être moniste et réaliste à l'égard de l'Arbre du vivant : l'histoire réelle du monde vivant devrait ainsi pouvoir être représentée par l'entremise d'un seul schéma arborescent. L'arbre est donc souvent considéré comme étant plus qu'un modèle utile (bien qu'il soit aussi considéré comme tel). De surcroît, le fait que le modèle arborescent soutende un principe ontologique suggère qu'une partie significative des phylogénéticiens et des phylogénéticiennes adopte une posture fondationnaliste à l'égard des connaissances développées dans leur champ d'études, une posture parfaitement illustrée par la philosophie de la preuve intégrale au sein de laquelle il est suggéré que les données « brutes » et nombreuses convergent et que cette convergence témoigne du fait que l'histoire évolutive est unitaire.

En plus d'offrir une analyse épistémologique détaillée de la phylogénétique traditionnelle, les pages qui précèdent permettent également de mettre en lumière un fait social important : la phylogénétique traditionnelle est ancrée dans une philosophie des sciences qui véhicule un empirisme naïf et dont les faiblesses ont été démontrées à maintes reprises dans les dernières décennies (Chang 2007; Lloyd 1995; Quine 1961; Sellars 1956; van Fraassen 1980, 2008). Afin de favoriser le bon fonctionnement des échanges au sein de cette discipline, pour s'assurer que ceux-ci ne soient pas ralentis par le réalisme et le monisme entretenus à l'égard du modèle arborescent (et le fondationnalisme sous-jacent) et dans l'espoir d'éviter que les prochains modèles (les modèles réseaux, par exemple) ne jouissent d'un traitement similaire, un regard épistémologique taillé sur des développements récents en philosophie des sciences doit être mis de l'avant. La charge théorique des observations, et plus largement l'ensemble des conséquences épistémiques du problème de la coordination, doivent être prises en compte au sein des débats scientifiques. Ces débats portent sur notre façon d'interagir avec des phénomènes, sur l'efficacité de ces

interactions et non sur la nature des phénomènes telle qu'elle existe (peut-être, qui sait?) dans un monde privé d'êtres humains, de leurs cultures intellectuelles et des théories qui les peuplent.

6.4 Le placement phylogénétique de *Methanonatronarchaeia*

Dans ce qui précède, j'ai offert une analyse critique de la phylogénétique traditionnelle centrée sur deux textes s'inscrivant à un haut niveau d'abstraction (Delsuc, Brinkmann et Philippe 2005; O'Hara 1997). Or, cette analyse prête le flanc à au moins deux critiques : d'abord, mon analyse est basée sur des textes qui datent. Ensuite, elle est construite sur des textes théoriques et non sur des textes de nature plus pratique. Afin de remédier à ces deux lacunes, je complète le chapitre par l'analyse d'un débat concret en biologie de l'évolution, un débat qui prend de surcroît la forme d'un échange tout à fait caractéristique de la phylogénétique traditionnelle, moniste et réaliste. Dans une série d'articles, publiés notamment dans la prestigieuse revue *Nature Microbiology* (Aouad et al. 2018; 2019; Sorokin et al. 2017, 2019), deux groupes de chercheurs et de chercheuses débattent du placement d'un phylum d'archées récemment découvert, soit *Methanonatronarchaeia*. Dans les pages qui suivent, je m'intéresse à la modulation de la structure inférentielle de la phylogénétique traditionnelle par chacun de ces groupes de chercheurs et chercheuses.

6.4.1 Le débat

Le débat concerne le placement d'un phylum d'archées récemment découvert appelé *Methanonatronarchaeia* (Sorokin et al. 2017). Les membres de ce phylum sont méthanogènes (producteurs de méthane par l'entremise de leur métabolisme) et halophiles (adaptés à la survie dans les milieux à haute salinité). Pour les chercheurs et chercheuses qui s'y intéressent, ce clade pourrait donc aider à comprendre l'évolution de ces deux traits très intrigants, le premier détenant une importance écologique marquée (le méthane étant un gaz à effet de serre) et le second témoignant de phénomènes adaptatifs à des conditions extrêmes. Or, comme les biologistes de l'évolution le réitèrent régulièrement, une compréhension de tels phénomènes gagne à être bâtie sur une connaissance phylogénétique du groupe concerné.

Un débat entre deux groupes de chercheurs et de chercheuses s'est déroulé à la suite de la publication originale, en 2017, de la découverte du phylum. Pour Sorokin et al., le groupe ayant identifié *Methanonatronarchaeia*, ce phylum serait un group frère de *Halobacteria*, phylum lui aussi caractérisé par l'halophilie (Sorokin et al. 2017, 2018, 2019). Pour Aouad et al., *Methanonatronarchaeia* serait à placer beaucoup plus profondément dans la phylogénie archée, à la base du groupe plus inclusif appelé *Methanotecta* (Aouad et al. 2018, 2019). Au-delà du placement du phylum en question, ce qui m'intéresse dans ce débat est la modulation du processus inférentiel par les groupes de chercheurs et de chercheuses impliqué-e-s. Il convient donc d'explorer le détail du débat.

Sorokin et al. ont formulé une première hypothèse phylogénétique concernant le placement de *Methanonatronarchaeia* en mobilisant deux jeux de données. D'abord, une analyse a été réalisée sur la base d'un alignement (réalisé avec le programme MUSCLE) de 51 protéines concaténées (les séquences concaténées contiennent plus de 8 000 nucléotides) pour 287 espèces, incluant les deux espèces de *Methanonatronarchaeia* étudiées dans cet article et un grand nombre d'archées (Sorokin et al. 2017). Les protéines utilisées ont été choisies parce qu'elles sont bien conservées chez les archées (Sorokin et al. 2017; Yutin et al. 2012). La super matrice résultant de cette analyse de similarité a par la suite été utilisée pour effectuer des inférences phylogénétiques, c'est-à-dire pour reconstruire un arbre phylogénétique des archées incluant le phylum *Methanonatronarchaeia* en utilisant une méthode statistique, soit le maximum de vraisemblance mis en œuvre par le programme PhyML3.0 (Guindon et al. 2010). Les résultats suggèrent que *Methanonatronarchaeia* serait un groupe frère de *Halobacteria*. Le tout a été confirmé par une analyse phylogénétique basée sur le marqueur génétique 16s r(RNA) et par le fait que les résultats convergent sur les topologies des archées généralement acceptées. L'une des implications de ces résultats est que l'halophilie aurait évolué une seule fois chez un ancêtre commun à *Methanonatronarchaeia* et *Halobacteria* (Sorokin et al. 2017).

En contraste, Aouad et al. (2019) ont proposé que la divergence de *Methanonatronarchaeia* soit beaucoup plus ancienne que Sorokin et al. ne le suggèrent, ce qui renforce l'hypothèse selon laquelle l'halophilie aurait évolué plusieurs fois indépendamment au sein des archées (deux fois selon Aouad et al. 2018; *Methanonatronarchaeia* serait donc, selon

ces chercheurs et chercheuses, le troisième phylum à acquérir indépendamment la capacité à survivre en milieu à haute salinité). Aouad et al. utilisent le même point de départ que Sorokin et al., soit la super matrice générée par la concaténation de 51 séquences protéiques. Cependant, ils appliquent un traitement supplémentaire à cette matrice en retirant les traits (chaque trait correspond ici à une colonne de l'alignement et donc à un nucléotide) qui évoluent plus rapidement, c'est-à-dire qui ont été sujet à de nombreuses substitutions nucléotidiques au travers du temps.

De manière générale, les sites évoluant rapidement sont considérés comme étant responsables d'un artéfact bien connu en phylogénétique, soit celui de l'attraction des longues branches. Les modèles utilisés pour l'inférence génétique, comme décrit au chapitre précédent, considèrent les modifications subies par des séquences génétiques comme des phénomènes stochastiques. Ainsi, lorsqu'un certain nombre de substitutions ont lieu sur un site donné, le signal phylogénétique se perd, en ce sens que toute paire de séquences a la même probabilité (0,25 si les substitutions sont purement aléatoires) de présenter le même nucléotide à cet endroit, que les phylums concernés soient phylogénétiquement proches ou non.

Sur le plan de la reconstruction phylogénétique, ce phénomène implique que les séquences détenant beaucoup de sites évoluant rapidement vont avoir tendance à se ressembler et donc à être groupées ensemble même si leur lien phylogénétique ne correspond pas à cette similarité qui est plutôt le fruit du hasard (et du fait d'avoir des sites qui évoluent rapidement). Leur représentation phylogénétique est de surcroît caractérisée par de longues branches puisque le haut taux de substitution implique une grande distance évolutive par rapport aux autres séquences étudiées. Le groupement de séquences évoluant rapidement et associées à de longues branches au sein de topologies phylogénétiques est donc considéré comme un artéfact à éviter (Delsuc, Brinkmann et Philippe 2005; Kostka et al. 2008; Sorokin et al. 2018).

Dans le cas de *Methanonatronarchaeia*, Aouad et al. considèrent que son association avec *Halobacteria* par Sorokin et al. résulte d'un tel artéfact. Ainsi, Aouad et al. utilisent la méthode *Slow-Fast*, telle qu'implantée dans le programme SLOW-FASTER (Kostka et al. 2008), afin d'extraire un signal phylogénétique plus net des séquences d'archées étudiées. Cette méthode,

reconnue comme une manière efficace d'identifier le phénomène d'attraction des longues branches et d'en minimiser l'impact sur la reconstruction phylogénétique (Brinkmann et Philippe 1999; Delsuc, Brinkmann et Philippe 2005), repose sur l'évaluation du taux de substitution des sites. Cette évaluation est basée sur l'étude des lignées identifiées de façon apriorique au sein de l'échantillon à l'étude comme étant monophylétique et permet d'écarter les sites dont le taux de substitution est plus élevé qu'un seuil choisi. L'étape est répétée à quelques reprises avec des seuils successivement plus bas, ce qui permet de faire des alignements qui négligent toujours plus de sites. Les topologies phylogénétiques résultant de chaque étape peuvent alors être évaluées pour déterminer si les suspicions d'attraction de longues branches sont fondées ou non. Une telle évaluation requiert une hypothèse phylogénétique par rapport à laquelle on peut comparer celles établies après l'utilisation de la méthode *Slow-Fast*, sans quoi aucun critère pour discerner la meilleure topologie (et donc l'existence d'artéfact dans certaines des autres hypothèses) ne serait disponible (Kostka et al. 2008).

Après l'utilisation de cette méthode et des inférences phylogénétiques réalisées à la fois par maximum de vraisemblance et méthodes bayésiennes, Aouad et al. obtiennent des résultats qui s'écartent significativement de ceux de Sorokin et al. Afin de soutenir leur hypothèse, Aouad et al. explorent cependant plus que la super matrice de Sorokin et al. dénuée des sites évoluant rapidement; ils et elles replacent les séquences associées à *Methanonatronarchaeia* au sein d'échantillons taxonomiquement plus fournis. Le phylum a été étudié en mobilisant un ensemble plus large de séquences (plus de 250 séquences protéiques formant une super matrice publiée préalablement dans Aouad et al. 2018). Il a aussi été appréhendé par l'entremise d'un échantillon de taxa plus inclusif (l'échantillon en question est aussi issu de travaux réalisés préalablement; Borrel et al. 2019). Dans les deux cas, le placement plus profond de *Metanonatronarchaeia* a été corroboré, de sorte que ces travaux rejettent l'hypothèse selon laquelle le phylum à l'étude serait un groupe frère d'*Halobacteria*.

L'histoire continue plus tard la même année alors que Sorokin et al. (2019) répondent directement à leurs interlocuteurs et interlocutrices. Leur réponse consiste en trois arguments qui portent sur les procédures phylogénétiques. Premièrement, Sorokin et al. affirment que la suppression des sites évoluant rapidement, bien qu'elle puisse effectivement minimiser l'impact

de signaux confondants, mène nécessairement à une perte d'information phylogénétique : « Removal of fast-evolving sites is a double-edged sword: it reduces the noise introduced by multiple substitutions, but phylogenetic information that is contained in comparatively variable positions is lost as well. » (Sorokin et al. 2019, 560; voir aussi Wägele et Mayer 2007). Ainsi, une reconstruction phylogénétique excluant les sites évoluant rapidement (ou ceux considérés comme tels) peut mener à une perte d'information importante. Au contraire, nous disent Sorokin et al., les sites bien conservés tendent à présenter peu de variation et donc peu d'information phylogénétique, la différence étant source d'information. En se basant exclusivement sur de tels sites bien conservés, il est donc possible de passer outre des relations de similarité significative, comme celle proposée par ces auteurs et autrices entre *Methanonatronarchaeia* et *Halobacteria*. Les liens de similarité entre ces deux phylums seraient dilués au sein d'une similarité générale et artificielle résultant de l'exclusion des sites les plus informatifs.

Deuxièmement, Sorokin et al. remettent en question l'utilité des matrices plus inclusives mobilisées par Aouad et al. pour soutenir leur hypothèse. Sorokin et al. signalent en effet que ces matrices sont composées de séquences protéiques codées par des gènes plus susceptibles d'être transférés horizontalement. Ainsi, l'étude de la similarité de ces séquences pourrait offrir des signaux au sujet des lignées de gènes qui ont des histoires phylogénétiques différentes de celle des lignées d'organismes qui les portent. Selon Sorokin et al., de telles séquences et les super matrices associées sont donc des sources d'information dont l'utilité doit être nuancée, d'autant plus que le marqueur génétique 16s r(RNA), dont l'utilité est quant à elle bien ancrée dans la discipline, suggère la sororité des lignées *Methanonatronarchaeia* et *Halobacteria*.

Troisièmement, une analyse qualitative des gènes partagés par les deux phylums à l'étude renforce, selon Sorokin et al., l'hypothèse selon laquelle il s'agit de groupes frères. Ils identifient des gènes codant pour des traits directement liés à l'halophilie et très peu présents chez les autres archées. La famille de gènes UspA liée à la réponse au stress est digne d'un intérêt particulier puisqu'elle est très développée au sein des deux phylums concernés et qu'elle est impliquée au sein de fonctions directement liées à l'halophilie. De plus, des gènes viraux associés à *Halobacteria* ont été observés chez *Methanonatronarchaeia*. Considérant que les virus interagissant avec des archées ont une gamme réduite d'hôtes possibles (Prangishvili et al. 2017),

les interactions virales communes aux deux phylums suggèrent une histoire évolutive partagée (Sorokin et al. 2019).

Le texte de Sorokin et al. évite cependant d'affirmer que l'hypothèse selon laquelle *Halobacteria* et *Methanonatronarchaeia* sont des groupes frères est la réponse définitive au problème. Il s'agit surtout, pour ces auteurs et autrices, de nuancer la portée de la reconstruction phylogénétique d'Aouad et al. Leur texte se conclut sur un appel à l'accumulation de plus de données au sujet de *Methanonatronarchaeia*, ce qui devrait pouvoir, selon eux et elles, mener le débat à bon port. Il convient de noter que depuis cet échange direct entre Aouad et al. et Sorokin et al., des recherches ont renforcé l'hypothèse du placement profond de *Methanonatronarchaeia* (Feng et al. 2021; Martijn et al. 2020). Aucune de ces recherches ne répond directement aux trois arguments de Sorokin et al. publiés en 2019 (Martijn et al. 2020 répond partiellement aux deux premiers arguments).

6.4.2 Analyse critique du débat

Le débat entre Sorokin et al. et Aouad et al. concernant le placement du phylum d'archées *Methanonatronarchaeia* récemment découvert confirme que l'analyse critique effectuée préalablement, centrée sur les textes d'O'Hara et de Delsuc et collègues, s'applique à certaines pratiques phylogénétiques contemporaines. Pour démontrer la chose, je me base sur les trois constats relatifs à la pensée arborescente détaillés précédemment.

Le premier constat critique que j'ai mis de l'avant vise à souligner que la pensée arborescente, associée à la phylogénétique traditionnelle, est composée d'une série d'amalgames. Premièrement, le modèle arborescent y est présenté comme étant diamétralement opposé à la pensée du groupe et à la pensée du développement. Cet amalgame (entre le modèle arborescent et une perspective opposée à ces deux schèmes de pensée) est bel et bien présent dans les travaux des groupes de chercheurs et de chercheuses étudiés, quoiqu'implicitement. Effectivement, il n'y a, dans les éléments théoriques et empiriques pertinents au débat, aucune raison de remettre en question cette association d'idées, de sorte qu'il soit tout à fait raisonnable d'éviter la répétition de principes solidement ancrés dans la discipline.

L'analyse critique du débat devient plus intéressante avec le deuxième amalgame, soit celui qui lie de façon nécessaire le modèle arborescent et l'évolution par voie de sélection naturelle. La phylogénétique, comme je l'explore en détail dans le premier chapitre de cette thèse, est une discipline qui permet notamment de fournir un contenu empirique à la théorie de l'évolution par voie de sélection naturelle. Lorsque le modèle arborescent est mobilisé pour ce faire, c'est non seulement la théorie de l'évolution qui se voit fournir un contenu empirique, mais aussi, plus spécifiquement, le postulat de la préséance des trajectoires divergentes au sein de l'histoire évolutive du monde vivant. Or, le débat concernant *Methanonatronarchaeia* témoigne d'au moins deux enjeux qui pourraient, si une ouverture logique sur ce plan était concédée, engendrer une remise en question de la préséance du patron arborescent. Ces deux enjeux concernent la possibilité de convergence évolutive et la présence de transferts génétiques horizontaux.

En ce qui concerne la convergence (je traite des transferts génétiques horizontaux plus loin), il faut surtout s'intéresser aux travaux d'Aouad et al. Dans leur série d'articles sur le sujet (Aouad et al. 2018, 2019; Borrel et al. 2019), ces biologistes affirment que la convergence est importante pour expliquer l'évolution de l'halophilie. Nonobstant ce constat central à leur propos, la convergence n'est représentée à aucun moment au sein des modèles inférentiels ou des figures qui représentent leurs résultats. Bien que la convergence soit reconnue comme un phénomène important pour comprendre l'évolution, nous disent Aouad et al., l'histoire évolutive des entités biologiques peut en faire abstraction, ne l'inclure ni au sein des figures qui témoignent des reconstructions phylogénétiques ni au sein des modèles inférentiels sous-jacents. La convergence est perçue comme un obstacle sur la route des inférences phylogénétiques, sur la découverte de l'histoire évolutive dont elle fait pourtant bel et bien partie. Les causes de la convergence, tout particulièrement dans le cas de l'halophilie, pourraient pourtant se révéler d'une importance capitale pour comprendre les phénomènes évolutifs concernés. Le fait que le modèle arborescent appliqué à la phylogénétique exclut ce type de relations témoigne ainsi d'une dissonance marquée au sein même de discours scientifiques le mettant de l'avant, comme celui d'Aouad et al. (2018, 2019).

Cette discussion nous amène directement au dernier amalgame, soit celui qui implique un lien de nécessité entre étude des lignées d'entités et étude de l'évolution. Tout le débat entourant le placement phylogénétique de *Methanonatronarchaeia* est ancré dans la validité de cette association d'idées. Comprendre les relations phylogénétiques de ce phylum devrait nous aider à mieux en comprendre l'évolution. Encore une fois, la convergence permet de douter, sinon de la chose, du moins de sa préséance au sein de la biologie de l'évolution. Considérons les différents liens de similarité qui unissent *Methanonatronarchaeia* et *Halobacteria* : similarités au niveau du gène 16s r(RNA) et de l'étendue de la famille de gènes UspA, observée avant exclusion des sites évoluant (potentiellement) rapidement, au niveau des gènes viraux présents dans les génomes, aux niveaux physiologiques et écologiques. Ce qui s'impose face à de tels constats est que la similarité qui lie ces deux phylums, qu'elle soit phylogénétiquement porteuse ou non, est d'une importance cruciale pour comprendre les phénomènes évolutifs. S'il s'agit effectivement de phénomènes mieux décrits par la convergence que la divergence de lignées, alors il pourrait effectivement s'agir de phénomènes évolutifs orthogonaux à la formation de lignées (les lignées convergent, mais ce n'est pas le processus de formation de lignées qui explique l'apparition des traits d'intérêt). Un cadre de pensée limité à l'étude des lignées, comme la phylogénétique, est donc potentiellement accessoire à l'étude de tels phénomènes évolutifs. Il est important de noter que ni Aouad et al. ni Sorokin et al. ne nient l'importance de l'ensemble des liens de similarité tout juste mentionnés. Cependant, les deux groupes de chercheurs et de chercheuses considèrent que le placement phylogénétique de *Methanonatronarchaeia* est important pour comprendre ces liens de similarité et leur portée évolutive, une suggestion s'effritant face à l'analyse critique du complexe discursif à l'étude.

Le deuxième constat critique énoncé plus tôt à l'égard de la pensée arborescente souligne que la pensée arborescente repose sur un principe ontologique, soit le principe d'histoire évolutive unique. Le modèle arborescent est ainsi considéré non seulement comme une représentation de connaissances phylogénétiques, mais aussi comme une structure réelle et unique. Les relations entre espèces ou autres unités taxonomiques seraient ainsi découvertes alors que diverses régions de l'arbre sont mises à jour.

Le débat étudié incarne bien ce principe ontologique : les auteurs et autrices impliqué-e-s cherchent à départager les hypothèses disponibles relativement au placement phylogénétique de *Methanonatronarchaeia*, rejetant implicitement la possibilité que l'histoire évolutive soit multiple. Cette posture est surprenante : les résultats des deux groupes de chercheurs et de chercheuses mettent en exergue le fait que les séquences moléculaires étudiées ont des histoires distinctes les unes des autres, et dans certains cas, distinctes de celle des organismes. Or, cette invitation à adopter une posture pluraliste est écartée *de facto* au profit d'une logique typique de la phylogénétique traditionnelle : l'histoire des gènes et celle des organismes qui les portent devraient converger, traduire les mêmes processus évolutifs, la même histoire. Ainsi, si un gène a une histoire qui s'écarte de celle des organismes qui le portent, l'information phylogénétique associée est considérée comme étant nuisible, comme du bruit (Bapteste et Dupré 2013). La phrase suivante en témoigne : « To address this issue, we performed an in-depth phylogenomic analysis designed to limit the impact of the non-phylogenetic signal on phylogenetic inferences. » (Aouad et al. 2018, 47, italiques ajoutés). Face à l'omniprésence des transferts génétiques entre lignées d'organismes, notamment chez les archées, et l'importance évolutive de ce phénomène (Bapteste et Boucher 2008; Doolittle 1999; McDaniel et al. 2010), cette posture semble d'autant plus problématique : tout porte à croire qu'une multitude d'histoires évolutives doit être racontée, soit celles de chaque séquence génétique impliquée et celle des lignées d'organismes³⁷. Les connaissances théoriques à l'égard de l'évolution, tout comme les données et le traitement qu'en font les auteurs et autrices concerné-e-s, suggèrent que ces histoires ne convergent pas et donc que leur représentation sous forme d'un seul arbre est un artéfact issu de l'adhérence des praticiens et des praticiennes à une croyance non justifiée et non justifiable. Ainsi, le principe ontologique d'histoire évolutive unique est conservé, ce qui témoigne de la force du principe de respect et de son rôle au sein du dynamisme scientifique qui caractérise la phylogénétique.

³⁷ La profondeur du problème pourrait être démontrée de manière encore plus forte en soulignant que de nombreux gènes sont des chimères ou composites, c'est-à-dire que ce sont des séquences génétiques formées de segments détenant des histoires évolutives distinctes (Méheust et al. 2018; Toleman et al. 2012; Watson, Lopez et Bapteste 2021). Ainsi, même l'histoire des gènes ne peut résumer l'histoire des séquences génétiques; à tous les niveaux, il y a une pluralité d'histoires à raconter.

Cette discussion nous amène finalement au troisième constat critique. En effet, le conflit entre le principe ontologique tout juste mentionné et les données démontre bien qu'afin de comprendre les dynamiques scientifiques, une approche philosophique qui, contrairement au cohérentisme historique, considère les données comme des fondations à la connaissance ou le facteur principal déterminant l'évolution des pratiques scientifiques est inadéquate. Dans le débat étudié, c'est un rapport mutuel entre données et contraintes théoriques qui assied les arguments et non une influence unidirectionnelle du « donné » ou de l'empirie vers les théories ou hypothèses scientifiques. Une négociation méthodologique influence entièrement ce qui peut être dit du monde empirique. Le tout est traduit par le débat au sujet du placement de *Methanonatronarchaeia* qui se révèle ultimement être d'ordre théorético-méthodologique (quelles données utiliser, comment procéder à leur curation, comment tenir compte [ou non] des transferts génétiques latéraux, de la convergence, des sites génétiques évoluant rapidement, etc.). Ce débat est ancré dans les itérations passées de la discipline et la légitimité qu'elles confèrent à certaines méthodes (par exemple, les programmes utilisés pour le traitement des données) et contraintes théoriques (par exemple, toutes celles liées au sein de la pensée arborescente).

Si cette situation semble évidente du point de vue épistémologique, les phylogénéticiens et phylogénéticiennes tendent tout de même à maintenir vivant le fondationnalisme comme posture épistémologique permettant de comprendre la phylogénétique ou même de la justifier. Alors même que ces auteurs et autrices sont impliqué-e-s dans un débat concernant principalement la méthodologie, ils et elles espèrent que l'accumulation de plus de données pourra régler le débat : « There is no doubt that more genomes will follow within a few years, which will likely provide the resolution of the current phylogenetic impasse. » (Sorokin et al. 2019, 561). La question se posera pourtant toujours, d'une itération à l'autre, obligeant les praticiens et praticiennes à négocier leur déchirement inévitable entre le principe de respect et l'impératif de progrès : les nouvelles données sont-elles recueillies, traitées et interprétées de façon adéquate? Affichant un retard malheureux par rapport aux constats épistémologiques issus de l'étude du problème de la coordination, la phylogénétique traditionnelle répond à cette question en présumant que la découverte des relations entre phylums offrira la réponse ultime à cet enjeu

théorético-méthodologique (comme si les données parlaient d'elles-mêmes). Cette inversion de la chaîne causale (les enjeux théorético-méthodologique précèdent la cueillette et le traitement de données et non l'inverse) témoigne du fait que la discipline baigne dans une culture intellectuelle fondationnaliste et qu'elle gagnerait à renouveler ses assises philosophiques afin de faciliter le dépassement des écueils des itérations passées. Une meilleure philosophie des sciences offre de meilleures descriptions des pratiques et permet ainsi de les manipuler de façon plus efficace (pour les rendre plus efficaces).

Avant de conclure, il convient de préciser que je mets l'accent, dans ce qui précède, sur les facteurs théoriques et méthodologiques influençant la pratique. Ces débats sont aussi sujets à des influences d'ordre sociohistoriques : le développement de nouveaux outils, l'accès à une plus grande puissance calculatoire, des découvertes connexes, l'organisation sociale de la communauté scientifique, etc. sont tout autant de facteurs dont une évaluation complète des pratiques scientifiques ne doit pas faire l'économie. En tant que philosophe, je me concentre cependant sur l'analyse qualitative d'un nombre limité de textes que j'ai tenté d'explorer en profondeur, mais cela ne doit aucunement être considéré comme un regard exclusif ou exhaustif sur l'objet d'étude, soit les pratiques phylogénétiques contemporaines.

6.4.3 Un bref retour sur le débat concernant *Methanonatronarchaeia*

Dans cette troisième partie du chapitre, j'ai présenté un débat phylogénétique récent qui porte sur le placement phylogénétique du phylum *Methanonatronarchaeia*. Ce débat m'a permis de démontrer que les constats critiques établis sur la base de l'analyse des textes d'O'Hara et de Delsuc et ses collègues sont toujours d'actualité. De plus, cela m'a permis de mettre en lumière le fait que les débats scientifiques, même lorsqu'ils sont présentés comme des débats portant sur des résultats ou une hypothèse, tendent à être centrés sur les modalités de la pratique qui mène à ces résultats. Ainsi, les outils du cohérentisme historique, une approche qui considère que le dynamisme scientifique est à saisir d'abord et avant tout par l'entremise des méthodes qui permettent dépasser les conséquences locales (au sein d'une discipline) du problème de la coordination, permettent d'analyser de tels débats. Au contraire, le fondationnalisme, qui considère que de tels débats sont ultimement déterminés par l'influence unidirectionnelle du

monde empirique sur les théories, modèles et hypothèses, occulte des éléments cruciaux du dynamisme scientifique. Les apories d'une telle approche sont explicites dans le débat étudié, alors que certains auteurs et autrices espèrent que l'accumulation de données pourra éventuellement clore un débat portant sur le traitement des données.

6.5 Conclusion

Dans ce chapitre, j'ai exploré la pensée arborescente et son rapport à la phylogénétique traditionnelle dont elle est le cadre théorique principal. J'ai aussi profité de l'occasion pour présenter le tournant génomique et son influence sur la phylogénétique. Pour ce faire, j'ai d'abord présenté un texte à saveur épistémologique (O'Hara 1997) qui brosse le portrait de la pensée arborescente et de ce qu'elle implique pour la phylogénétique de la fin du 20^e siècle. J'ai ensuite présenté un texte programmatique s'intéressant à la phylogénomique (Delsuc, Brinkmann et Philippe 2005) qui illustre bien la mise en œuvre de la pensée arborescente.

Trois constats découlent de l'analyse de ces deux textes. Premièrement, la pensée populationnelle est un cadre de pensée beaucoup plus complexe que ce qui est suggéré par la description qu'O'Hara en donne. Autrement dit, les idées et les relations qu'elles entretiennent sont plus nombreuses et variées que ce que le texte de 1997 le laisse entendre. Spécifiquement, j'ai identifié trois amalgames qui gagnent à être décortiqués pour le bien des descriptions épistémologiques de la phylogénétique et pour favoriser les débats de qualité au sein de la discipline scientifique concernée. Le premier amalgame consiste en l'association entre le modèle arborescent, d'une part, et l'opposition aux pensées du groupe et du développement, de l'autre. Contrairement à ce qu'O'Hara suggère, le modèle arborescent est tout à fait compatible avec la pensée du groupe et celle du développement. Il est vrai que l'utilisation du modèle arborescent contemporaine se dissocie de ces deux schèmes de pensée, mais cela ne justifie aucunement l'amalgame véhiculé par le texte d'O'Hara.

Le deuxième amalgame revient à lier évolution par voie de sélection naturelle et modèle arborescent. Je suis loin d'être le premier à remettre en question la nécessité de cette association d'idées. La phylogénétique basée sur les réseaux, par exemple, émane d'une réévaluation de la

capacité du modèle arborescent à rendre compte adéquatement des phénomènes d'évolution. Je reviens en détail sur ce sujet dans le prochain chapitre.

Le troisième amalgame consiste à lier de façon nécessaire l'étude de l'évolution et l'étude de lignées. L'évolution par voie de sélection naturelle est un phénomène au sein duquel des lignées peuvent figurer et a, depuis sa première théorisation, été étudiée sur la base de cette possibilité. Il va sans dire que ce programme de recherche a été fructueux et qu'il le sera encore. Cependant, ces succès n'impliquent aucunement un lien nécessaire entre évolution par voie de sélection naturelle et étude de lignées biologiques puisque ce type d'évolution peut effectivement être réalisé sans que des lignées soient formées. Dans ce chapitre, j'ai résumé un argument allant en ce sens qui a été publié ailleurs (Papale 2021).

Après avoir mis en lumière cette série d'amalgames, le deuxième constat tiré de l'analyse de la phylogénétique traditionnelle est que le modèle arborescent y agit comme un principe ontologique, c'est-à-dire une croyance considérée comme valide et utilisée pour structurer la pratique sans être justifiée. C'est plutôt la place occupée par le principe de respect au sein de la systématique qui confère sa centralité au modèle arborescent, et ce, depuis l'époque prédarwinienne. Le fait que le modèle arborescent puisse être théorisé comme étant un principe ontologique implique une posture réaliste et moniste à son égard qui est largement partagée au sein de la phylogénétique.

Ce point nous amène au troisième constat, soit que les phylogénéticiens et phylogénéticiennes adhèrent implicitement et parfois explicitement (comme dans le cas de la philosophie de la preuve intégrale) à une philosophie fondationnaliste. Cela implique notamment qu'ils et elles considèrent que les données sont les seuls couteaux permettant de trancher les débats scientifiques. À l'aune du problème de la coordination, une telle posture philosophique apparaît comme étant indéfendable : l'observation et le traitement des données sont des gestes épistémiques théoriquement chargés et ne peuvent être réalisés qu'une fois que des procédures et contraintes théoriques sont acceptées. Les débats ne peuvent donc être tranchés (provisoirement) que par une réflexion sur ces procédures ou contraintes (à l'exception de débats

pour lesquels les procédures d'observation sont complètement stabilisées, ce qui est loin d'être le cas en phylogénétique).

Le débat étudié dans la troisième partie du chapitre a permis de confirmer la validité des trois constats tout juste résumés, mais il a surtout mis en exergue les difficultés associées au fondationnalisme. Ce débat, présenté par les auteurs et autrices impliqués·e·s comme étant relatif au placement phylogénétique de *Methanonatronarchaeia*, est plus rigoureusement décrit comme un débat sur les modalités d'observation des liens phylogénétiques entre phylums. Les deux groupes de chercheurs et de chercheuses (Aouad et al. 2019; Sorokin et al. 2018) adhèrent à la pensée arborescente, mais ils négocient différemment les modalités d'application des principes théoriques plus abstraits qu'elle impose. La possibilité même de tels échanges découle du fait que le phénomène étudié, soit les liens phylogénétiques entre unités taxonomiques, est inobservable et que les méthodes permettant son observation indirecte sont sous-déterminées par les phénomènes et non stabilisées. Le fondationnalisme, puisqu'il balaie cette réalité sous le tapis de l'empirisme naïf comme je l'ai démontré au chapitre 1, est incapable d'encadrer épistémologiquement et adéquatement les pratiques scientifiques ainsi soumises au problème de la coordination.

Chapitre 7 – De la comparaison entre deux modèles phylogénétiques : les réseaux et les arbres

Ne t'excuse jamais de faire de la philo.

- Frédéric Bouchard,
lors d'une rencontre de travail

7.1 Introduction

L'objectif principal de ce chapitre est de réaliser une analyse comparative des pratiques phylogénétiques centrées sur les réseaux et de celles centrées sur le modèle arborescent. Pour ce faire, je présente d'abord en détail les pratiques phylogénétiques axées sur les réseaux. Cet ensemble de pratiques étant hautement hétérogène, j'appréhende cette diversité par l'entremise de deux dichotomies : 1) les réseaux peuvent être abstraits ou explicites et 2) ils peuvent être enracinés ou non. En 7.2.1, je présente ces quatre qualificatifs et donne des exemples concrets pour les illustrer. En 7.2.2, je présente l'histoire des modèles réseaux en biologie, avant de réaliser en 7.2.3 l'analyse comparative des deux itérations de la phylogénétique étudiées. Le tableau 7.1 présente les résultats de cette analyse de façon synoptique. Il ressort de l'analyse que le rapport entre modèles réseaux et modèles arborescents, en phylogénétique, est déterminé par une négociation entre le principe de respect et l'impératif de progrès : les modèles réseaux offrent la possibilité d'inférer l'existence de phénomènes évolutifs importants, soit ceux qui mènent à la réticulation des lignées, sur la base de données que la phylogénétique traditionnelle considère être du bruit (c'est-à-dire qu'elle écarte ces données « conflictuelles » pour réaliser ses inférences); ce progrès interne de la discipline, résultant de la capacité à rendre observable ce qui ne l'est pas directement, s'assied toutefois sur une structure inférentielle qui demeure largement intacte. Ainsi, la phylogénétique centrée sur les réseaux ne s'aventure dans de nouvelles contrées que si celles-ci sont à proximité de la phylogénétique traditionnelle, cette itération passée dont elle tire sa légitimité.

Dans la deuxième partie du chapitre (7.3), j'analyse cinq débats philosophiques directement liés aux pratiques phylogénétiques par rapport auxquels ma posture est informée

par l'analyse comparative réalisée préalablement. Les trois premiers débats sont présents dans le corpus de textes étudiés dans cette thèse. Il s'agit premièrement d'évaluer l'incompatibilité des modèles réseaux avec le cadre darwinien telle que suggérée par certains et certaines biologistes. Je démontre qu'une telle incompatibilité est artéfactuelle puisqu'elle découle d'analyses épistémologiques qui manquent de précision. Deuxièmement, il est question d'adopter une posture quant à la teneur réaliste des deux types de modèles étudiés. En m'appuyant sur les écrits de Velasco (2012, 2013), je démontre que dans le contexte du réalisme, les modèles réseaux sont plus adéquats. Par contre, j'insiste surtout sur le fait qu'à la lumière du problème de la coordination, la question même du réalisme (et de l'antiréalisme) perd de son lustre. Troisièmement, j'appréhende la question de l'utilité respective des deux types de réseaux, suggérant alors une posture flexible et pragmatique. Le cohérentisme historique nous permettant de nous défaire de la question du réalisme par rapport aux modèles, il convient simplement d'utiliser ceux qui nous permettent le plus efficacement de répondre à nos objectifs de recherche. Les modèles arborescents, malgré leurs degrés de complexité et de précision moindre que ceux des modèles réseaux, ont toujours un rôle à jouer en phylogénétique.

Cette considération par rapport aux objectifs de la recherche mène tout naturellement aux deux derniers débats dont je traite dans ce chapitre. Ces deux débats sont absents du corpus de textes étudiés, mais méritent tout de même de recevoir l'attention des philosophes et des biologistes. Ils portent sur les objectifs de la phylogénétique et sur les modalités de réalisation de ces objectifs. Le quatrième débat porte spécifiquement sur le rapport entre l'étude des lignées et l'étude de l'évolution. J'ai démontré au chapitre précédent que le lien nécessaire entre formation de lignées et évolution par voie de sélection naturelle est le fruit d'un amalgame. Dans la section 7.3.4 du présent chapitre, j'évalue les implications de ce constat pour la phylogénétique, dont l'objectif est de retracer les liens généalogiques entre unités taxonomiques dans le but d'enrichir le contenu empirique des théories de l'évolution.

Enfin, le cinquième débat considéré porte sur la place de la phylogénétique, en tant que façon de décrire la biodiversité, au sein de la biologie et surtout de la biologie de l'évolution. Je souligne alors la nature classificatoire du projet phylogénétique qui prime sur sa dimension historique. La dimension classificatoire de la phylogénétique implique une lacune importante :

elle néglige les différences entre les membres d'une unité taxonomique, et de ce fait, elle entre en conflit avec la logique darwinienne qu'elle est censée incarner. En conséquence, je considère que la biologie de l'évolution gagnerait à être centrée sur des méthodes descriptives quantitatives³⁸. Loin de moi l'idée d'écarter la classification de la biologie de l'évolution; ma suggestion est simplement de faire primer d'autres modes descriptifs dans le cadre de notre rapport gnoséologique à la biodiversité.

7.2 Qu'est-ce qu'un réseau phylogénétique et en quoi cela diffère d'un arbre phylogénétique?

7.2.1 Les réseaux phylogénétiques et la théorie des graphes

Pour saisir ce qu'est un réseau phylogénétique, il faut d'abord comprendre ce à quoi le terme « réseau » renvoie plus largement. Les réseaux sont des outils et objets d'étude de la branche des mathématiques appelée théorie des graphes (Newman 2018). Les termes *graphe* et *réseau* peuvent théoriquement être considérés comme des synonymes, bien qu'en pratique, le terme graphe tend à être utilisé pour référer aux représentations mathématiques (et parfois exclusivement aux représentations de petite taille) des réseaux considérés quant à eux comme des systèmes concrets. Un réseau (j'adopte ici la synonymie liant les termes réseau et graphe) se définit comme un ensemble de composantes qui sont liées les unes aux autres. Le tout est représenté par l'entremise de nœuds (*nodes*) qui font office de composantes et d'arêtes (*edges*) qui illustrent les liens directs entre les composantes (Barabási 2016). En ce sens, les réseaux sont indéniablement des outils très versatiles capables de représenter des phénomènes variés (des systèmes de relations sociales, des écosystèmes, des systèmes hydrographiques, les relations généalogiques entre unités taxonomiques, des liens inférentiels entre concepts, etc.). La théorie des graphes vient ainsi étudier les propriétés topologiques des réseaux, comme le degré moyen

³⁸ Les méthodes descriptives mentionnées dépendent ultimement des descriptions qualitatives qui permettent de donner un sens aux abstractions nécessaires à la quantification ainsi qu'à la classification. L'étude des méthodes qualitatives, qui offrent les descriptions les plus détaillées des phénomènes, est un sujet qui brille par son absence au sein de la philosophie des sciences. Considérant l'étendue du sujet, je dois cependant faire l'économie de la chose dans ce contexte-ci.

des graphes, leur structure, les motifs qui les composent, leur taille, leur connectance, différentes propriétés des chemins liant les nœuds, etc. de sorte à mettre le doigt sur ce qui distingue les réseaux aléatoires des réseaux qui ne le sont pas. L'utilisation de la théorie des graphes pour appréhender des phénomènes concrets découle en effet de cette prémisse : les systèmes biologiques sont organisés de façon non aléatoire (Huneman 2010). Comprendre les propriétés topologiques de ces phénomènes devrait ainsi pouvoir nous aider à mieux saisir leurs particularités, leur fonctionnement.

Les arbres phylogénétiques sont des réseaux au sens de la théorie des graphes. Plus précisément, il s'agit de réseaux à structure bifurquée au sein desquels chaque nœud intérieur est lié à exactement trois autres nœuds (dans un arbre enraciné, ces trois nœuds représentent un taxon ascendant et deux taxons descendants). Les nœuds ou taxons terminaux sont liés exclusivement à leur ancêtre. Les arbres phylogénétiques enracinés (au sein desquels un ancêtre commun postulé qui fait office de racine) sont des réseaux dirigés (les arêtes ont une *direction*, soit celle du temps associé aux liens généalogiques entre taxons) et acycliques (aucun chemin partant d'un nœud et respectant la direction des arêtes ne peut revenir au nœud de départ) qui traduisent une hypothèse précise relativement à la séquence de divergence des taxons étudiés (Morrison 2005). Strictement parlant, une opposition entre arbres phylogénétiques et réseaux est donc maladroite. Cependant, dans le corpus étudié, le terme « réseau » est utilisé pour référer aux graphes qui incluent des liens transversaux entre les lignées et qui, en ce sens, se distinguent de la structure bifurquée des arbres phylogénétiques. À moins d'une précision contraire, j'utilise les deux termes (arbres et réseaux) en cohérence avec leur acception phylogénétique plutôt que leur signification mathématique.

Les réseaux phylogénétiques se caractérisent donc d'abord et avant tout par leur rejet de la structure arborescente au profit d'une représentation des liens plus complexes entre taxons. Mathématiquement parlant, les nœuds du réseau peuvent ainsi être liés à plus de trois autres nœuds et surtout à deux ancêtres (ou plus, bien que le nombre d'ancêtres soit généralement limité à deux; j'y reviens plus loin). Biologiquement parlant, l'objectif est de représenter des liens phylogénétiques dits horizontaux ou latéraux entre taxons (Baptiste et al. 2004; Baptiste et Boucher 2008; Doolittle 1999, 2000; Francis et Steel 2015; Huson et Scornavacca 2011; Huson,

Rupp et Scornavacca 2010; Morrison 2005; Than, Ruths et Nakhleh 2008). Ces liens correspondent à des phénomènes tels que l'hybridation de lignées d'organismes pluricellulaires (Mallet, Besansky et Hahn 2016; Ottenburghs 2019; Suarez-Gonzalez, Lexer et Cronk 2018), les transferts génétiques latéraux (ou horizontaux; Bapteste et al. 2004; Bapteste et Boucher 2008; Doolittle 1999; Keeling et Palmer 2008; McDaniel et al. 2010), les différents processus de recombinaison de matériel génétique (Méheust, Lopez et Bapteste 2015; Méheust et al. 2018; Naor et al. 2012; Toleman et al. 2012) ou l'endosymbiose (Lane et Archibald 2008; Martin, Garg et Zimorski 2015).

Malgré le désir commun de rendre compte de ces événements évolutifs qui engendrent de la réticulation au sein des liens phylogénétiques, l'utilisation de l'expression « réseau phylogénétique » réfère à un ensemble de pratiques hétérogène. Une exploration plus détaillée de la chose permet de cerner certaines spécificités des pratiques qui y sont associées. Deux distinctions importantes (qui s'appliquent aussi aux modèles arborescents) doivent être considérées. D'abord, il existe une distinction entre réseaux dits évolutifs (Morrison 2005; 2010) ou explicites (Huson, Rupp et Scornavacca 2010; Huson et Scornavacca 2011), d'un côté, et réseaux dits de visualisation des données (*data-display networks*) ou abstraits, de l'autre. Ensuite, il faut distinguer les réseaux enracinés des non enracinés.

La distinction entre réseaux explicites et abstraits est d'une importance capitale. Les réseaux explicites sont des représentations d'hypothèses phylogénétiques, c'est-à-dire que les relations ainsi illustrées sont considérées comme relayant l'histoire évolutive des taxons concernés (comme les arbres phylogénétiques enracinés). Au contraire, les réseaux abstraits représentent des relations entre taxons, gènes ou familles de gènes sans pour autant être des hypothèses phylogénétiques (Morrison 2005, 2010; Huson 2007; Huson et Scornavacca 2011). Il peut s'agir, par exemple, de mises en relation basées sur des études de similarité (comme des alignements de séquences) qui visent principalement à présenter les résultats de ces analyses sans associer directement ces résultats à des liens généalogiques. Jumelés aux outils d'analyse des théories des graphes, les réseaux abstraits ouvrent les portes d'une grande variété d'analyses des relations entre unités taxonomiques qui permettent autant des inférences fonctionnelles (Bernard et al. 2020; Watson et al. 2019), taxonomiques (Corel et al. 2018; Iranzo, Krupovic et Koonin 2016; Iranzo et al. 2016) que généalogiques. Or, au sens propre de l'adjectif

phylogénétique (« relatif à la phylogenèse », nous dit le Robert), ces réseaux abstraits ne sont pas nécessairement phylogénétiques. Il s'agit plus spécifiquement d'outils mathématiques permettant l'analyse de la similarité qui peuvent notamment être utilisés pour inférer l'histoire évolutive d'unités taxonomiques. Les réseaux explicites, quant à eux, sont indéniablement phylogénétiques en ce sens qu'ils traduisent (explicitement) la phylogenèse d'un ensemble d'unités taxonomiques. Avant d'illustrer cette distinction, il convient de présenter la seconde.

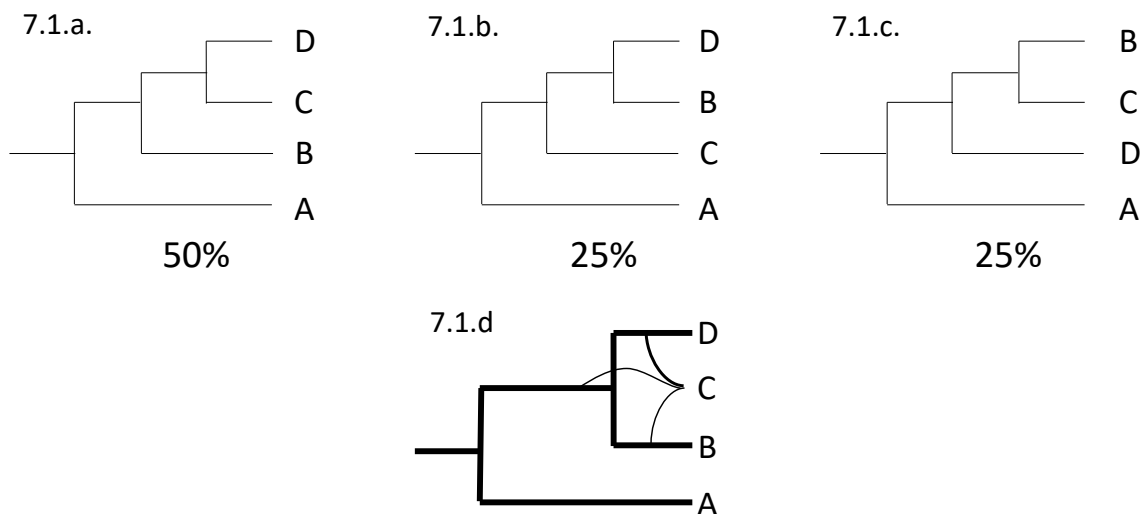
La distinction entre les réseaux enracinés et non enracinés rappelle celle entre les réseaux explicites et abstraits sans parfaitement s'y conformer (Huson, Rupp et Scornavacca 2010; Huson et Scornavacca 2011). Les réseaux enracinés, comme leurs contreparties arborescentes, sont dirigés et acycliques. Une origine évolutive de l'ensemble de taxons à l'étude, soit la racine, est postulée comme point de départ de l'histoire représentée par le réseau. Les relations généalogiques figurant dans le réseau émanent de cette racine et suivent la direction générale qui est celle du temps sans toutefois se limiter au patron bifurqué. Tout réseau explicite est un réseau enraciné. En effet, sans racine ou point de départ évolutif, il est impossible de représenter une séquence concrète de transformations évolutives des unités taxonomiques.

Cette dernière affirmation devient plus claire en considérant ce qu'est un réseau non enraciné. Il s'agit d'un ensemble de nœuds mis en relation sans qu'il y ait un point de départ prédéterminé. Ainsi, le réseau peut être parcouru par différents chemins commençant à différents endroits. Ces modèles ne sont donc pas dirigés, ce qui rend impossible leur interprétation directe en tant qu'hypothèses phylogénétiques. Pour les interpréter phylogénétiquement, il faut les enraciner, c'est-à-dire choisir un nœud comme étant l'ancêtre commun à l'ensemble des unités taxonomiques étudiées, un point zéro à partir duquel la séquence des transformations s'opère.

Il faut cependant éviter une association hâtive entre réseaux enracinés et explicites. En effet, des réseaux enracinés peuvent aussi être abstraits. C'est le cas des réseaux de groupes (*cluster networks*). Ceux-ci sont utilisés pour représenter en un seul graphe et de façon abstraite un ensemble d'arbres phylogénétiques enracinés qui présente des groupes incompatibles (*incompatible clusters*; Huson, Rupp et Scornavacca 2010). Les réseaux de groupes représentent

de façon unitaire ces différentes hypothèses phylogénétiques incompatibles afin d'identifier les potentiels groupes monophylétiques qui demeurent intacts même lorsque l'on tient compte des conflits entre les arbres sous-jacents (les groupes monophylétiques ainsi identifiés sont donc généralement des *clusters* plus inclusifs que ceux des arbres qui sont incompatibles les uns avec les autres). Ces réseaux sont (au moins partiellement) abstraits puisqu'ils traduisent une posture agnostique au sujet des séquences de transformations au sein des *clusters* où des hypothèses alternatives sont représentées (aucune explication phylogénétique des incohérences entre arbres n'est offerte). La Figure 7.1 (fortement inspirée de Huson et Scornavacca 2011, 31) illustre la chose.

Figure 7.1 – Un exemple de réseau phylogénétique enraciné mais abstrait



Cette figure illustre l'utilisation de réseaux phylogénétiques enracinés mais non explicites. Le réseau en 7.1.d représente l'information phylogénétique contenue dans 7.1.a, 7.1.b et 7.1.c sans privilégier l'une des hypothèses relatives aux relations entre espèces B, C et D et sans expliquer les incompatibilités. La largeur des traits en 7.1.d témoigne du soutien empirique de chaque hypothèse et correspond aux pourcentages indiqués en 7.1.a, 7.1.b et 7.1.c.

Il est cependant à noter que les réseaux de groupes peuvent être interprétés de façon explicite. Dans une telle optique, on peut calculer les réseaux phylogénétiques enracinés minimaux (*minimum rooted phylogenetic networks*), soit ceux qui impliquent le moins de réticulation possible. En effet, Huson et ses collègues nous disent la chose suivante (dans cette

citation, N réfère à un réseau donné et C à un ensemble de groupes [*clusters*] utilisés pour inférer le réseau N) :

In a biological interpretation of N , the reticulate nodes will ideally correspond to reticulate evolutionary events such as speciation by hybridization, horizontal gene transfer, or recombination of closely related sequences, depending on the context. If we assume that such evolutionary events occur only rarely, then this justifies that we require that a rooted phylogenetic network N representing C should contain as few reticulations as possible. (Huson, Rupp et Scornavacca 2010, 149)

Les réseaux phylogénétiques enracinés et explicites (on considère que la réticulation y représente des événements évolutifs concrets) ainsi inférés sont donc notamment structurés par la croyance selon laquelle la réticulation est l'exception et non la règle.

En résumé, la distinction entre réseaux abstraits et explicites et celle entre réseaux enracinés et non enracinés ne se chevauchent pas exactement. Dans ce qui suit, j'approfondis la première des deux, soit celle entre réseaux abstraits et explicites, qui fait écho à la distinction entre les rôles inférentiel et de représentation phylogénétique que joue le modèle arborescent. Cela me permet, en 7.2.3 et dans la deuxième partie de ce chapitre, de mener efficacement l'analyse comparative entre les modèles réseaux et les modèles arborescents.

7.2.1.1 Les réseaux abstraits

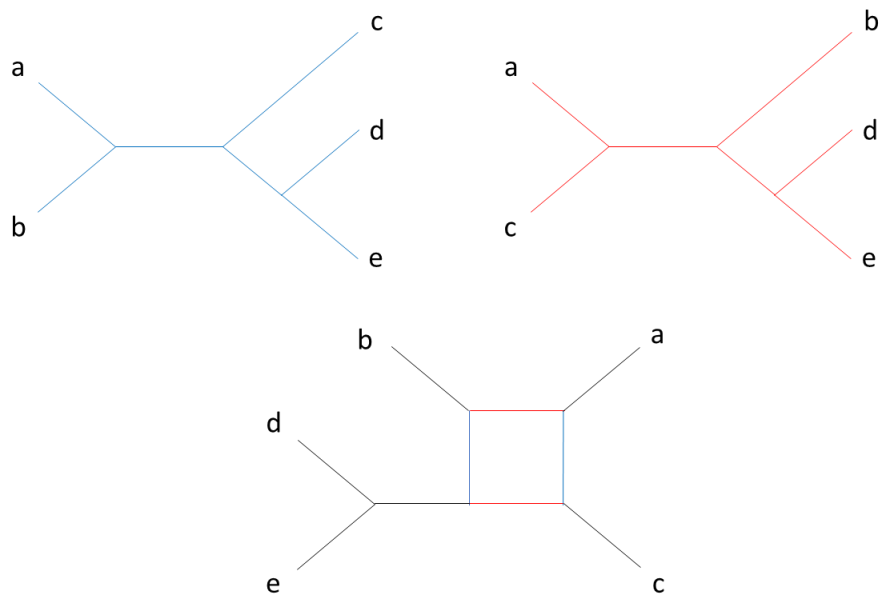
Les distinctions entre réseaux abstraits et explicites ainsi qu'entre réseaux enracinés et non enracinés étant présentées, je me penche maintenant sur la première de ces deux distinctions, qui est d'une grande importance pour comprendre la transition des modèles arborescents vers les modèles réseaux. Pour ce faire, j'illustre les deux types de réseaux tour à tour de sorte à les inscrire au sein de pratiques phylogénétiques concrètes.

Les réseaux de bipartitions (*split networks*)³⁹ sont les réseaux abstraits et non enracinés les plus couramment utilisés en phylogénétique (Huson et Scornavacca 2011; Morrison 2014b). Ces réseaux visent à représenter l'ensemble des bipartitions possibles pour un ensemble de données. Une bipartition est la division d'un ensemble en deux sous-ensembles. Ainsi, chaque

³⁹ À ne pas confondre avec les graphes bipartis.

embranchement d'un arbre (phylogénétique, par exemple) est une bipartition puisqu'il représente une décomposition d'un groupe en deux groupes moins inclusifs. Un jeu de données au sein duquel toutes les bipartitions (obtenues sur la base d'analyse de similarité) sont compatibles les unes avec les autres peut être représenté sous forme d'un arbre phylogénétique. Or, les arbres phylogénétiques ne représentent qu'une hypothèse phylogénétique à la fois, alors que dans certains cas, plusieurs bipartitions non compatibles sont disponibles pour un seul et même jeu de données. Les réseaux de bipartitions permettent de représenter les différentes bipartitions disponibles au sein d'un seul et même graphe, même si ces bipartitions sont incompatibles. La Figure 7.2 illustre la chose. De tels réseaux sont bel et bien abstraits puisqu'il n'est aucunement question de représenter une séquence de transformations phylogénétiques, une hypothèse concrète. Au contraire, ces réseaux rendent explicite le fait qu'un ensemble de données peut mener à deux topologies arborescentes distinctes.

Figure 7.2 – Un réseau de bipartitions et les deux arbres qu'il représente



Cette figure présente un réseau de bipartitions (7.2.c) et deux arbres (7.2.a et 7.2.b) issus de bipartitions incompatibles de l'ensemble {a, b, c, d, e}. Le dédoublement des arêtes en 7.2.c permet de représenter les deux bipartitions : celle dans laquelle a et b sont des groupes frères (en faisant abstraction des arêtes rouges, les deux arêtes bleues seraient fusionnées et l'arête ainsi engendrée représenterait la bipartition de 7.2.a) et celle dans laquelle a et c sont des groupes frères (en excluant les arêtes bleues, on obtient la bipartition de 7.2.b). Le code de couleurs est ici utilisé exclusivement pour faciliter la compréhension du graphe.

Dans un texte de 2005, Morrison met de tels réseaux en œuvre. Il revient sur une phylogénie de neuf haplotypes (un ensemble d'états de caractères généralement transmis en bloc d'une génération à l'autre) d'une espèce de mite parasitique (*Sarcoptes scabiei*) publiée par Skerratt et al. (2002) sur la base d'une étude de petites sous-unités de ARN ribosomal de mitochondrie. Skerratt et al. présentent une phylogénie arborescente des neuf haplotypes, mais celle-ci n'est que l'une des topologies qu'ils et elles ont pu inférer sur la base des données utilisées et leur traitement par maximum de parcimonie. Selon Morrison, cette situation témoigne du fait que le modèle arborescent ne convient pas au traitement des données : « the parsimony analysis assumes that the underlying model for the evolutionary history of the organisms is a dichotomously branching tree. It is highly questionable whether this is an appropriate model for the relationships between individuals within a species. » (Morrison 2005, 574).

Afin de corriger la situation, Morrison réanalyse le même jeu de données (en y ajoutant une séquence génétique issue de la même espèce, mais d'un échantillon obtenu en Suède plutôt qu'en Australie comme c'est le cas pour les échantillons de Skerratt et al.), notamment par l'étude des bipartitions incompatibles. Il considère que le réseau de bipartitions qu'il reconstruit comme une représentation complète des données utilisées (« a complete visual representation of the data » [Morrison 2005, 575]) offre un regard plus exhaustif sur les relations potentielles entre les haplotypes étudiés. Plus spécifiquement, une telle visualisation des données permet d'identifier plus facilement le lieu et la portée des incohérences phylogénétiques entre les caractères étudiés. Grâce à la visualisation du réseau de bipartitions, Morrison souligne que les séquences étudiées concordent toutes sur la même phylogénie (ou presque; il resterait tout de même de la réticulation dans l'arbre, mais rien d'important) si un caractère (un nucléotide dans chaque séquence) est retiré de l'analyse. Morrison affirme en effet la chose suivante :

Without this character, there would only be one reticulation in the network [...], and so it would be almost tree-like. The tree shown in Fig. 1a is thus misleading, as it indicates poor support for all of the branches but does not show that this lack of support is created by only one conflicting character. (Morrison 2005, 575)

Ainsi, les réseaux de bipartitions peuvent être utilisés pour analyser les données sans pour autant générer d'hypothèses généalogiques réticulées (il s'agit bien de réseaux abstraits). Dans le cas précis qui l'intéresse, Morrison n'offre aucune explication de la réticulation qui demeure

présente après avoir retiré le caractère générant l'incompatibilité; il n'offre aucune explication de l'incompatibilité observée sur ce caractère non plus. L'apport du modèle réseau est ici d'offrir un meilleur traitement des données, et plus précisément, des données qui témoignent d'incompatibilité au sein d'analyses arborescentes. Les réseaux sont d'abord et avant tout des outils mathématiques d'une grande puissance, cet exemple en témoigne bien. Une partie significative de l'utilisation d'outils issus de la théorie des graphes en phylogénétique (les réseaux de similarité de séquences, les réseaux bipartites, etc.) est à comprendre de la sorte : il s'agit d'outils qui facilitent le traitement et la représentation des données et non de modèles visant à représenter la réalité ou à expliquer les phénomènes.

7.2.1.2 Les réseaux explicites

Alors que certains outils utilisés pour l'inférence de réseaux explicites sont les mêmes que ceux utilisés pour l'inférence de réseaux abstraits, l'interprétation phylogénétique des deux types de réseaux demeure distincte. Les réseaux explicites, *de facto* enracinés, traduisent des hypothèses phylogénétiques, c'est-à-dire des hypothèses relatives à l'histoire évolutive des lignées étudiées. Un exemple, peut-être le plus fameux et influent du lot, est celui fourni par Doolittle (2000) dans son article « Uprooting the Tree of Life ». Doolittle y décrit d'abord la réorganisation progressive de l'Arbre du vivant par Woese (Woese 1987; Woese, Kandler et Wheelis 1990). Auparavant, l'Arbre du vivant était divisé en deux troncs principaux, soit celui représentant les procaryotes (le terme étant considéré à l'époque comme un synonyme de bactérie) et celui représentant les eucaryotes. Woese a démontré l'existence d'une divergence demandant de scinder les procaryotes en deux groupes, soit les bactéries et les archées, tôt dans l'histoire évolutive du monde vivant.

Toujours dans le même article, Doolittle (2000) synthétise par la suite les différents résultats des années 1980 et 1990 relatifs à la phylogénie du triptyque du monde vivant s'étant établie comme l'hypothèse dominante. Ces différents résultats convergent sur une hypothèse ayant été formulée préalablement, soit celle de l'endosymbiose : les eucaryotes auraient émergé d'une symbiose entre une lignée d'archées et des lignées de bactéries; ces dernières seraient à l'origine de la mitochondrie et des chloroplastes. Doolittle (2000) ajoute à ce phénomène d'endosymbiose les constats relatifs aux transferts génétiques latéraux, qui se font de plus en

plus nombreux à l'époque, avant de conclure que l'arbre du vivant est un modèle incapable de représenter adéquatement l'histoire évolutive des trois grands domaines du monde vivant :

[The] revised 'Tree' of life retains a treelike structure at the top of the eukaryotic domain and acknowledges that the eukaryotes obtained mitochondria and chloroplasts from bacteria. But it also includes an extensive network of untreelike links between branches. Those links have been inserted somewhat randomly to symbolize the rampant lateral gene transfer of single or multiple genes that has always occurred between unicellular organisms. (Doolittle 2000, 95)

La figure accompagnant cette citation est en effet truffée de réticulation, de sorte qu'il s'agisse bien d'un réseau et non d'un arbre. Il faut bien noter, comme en témoigne la citation de Doolittle, que dans ce cas-ci, le réseau est un modèle permettant la représentation et non l'inférence de liens phylogénétiques. La réticulation représentée a en effet été incluse dans la représentation de manière plutôt aléatoire (*somewhat randomly*) et non pour donner suite à des analyses réseaux. Nonobstant cette limitation, le réseau de Doolittle a eu une influence marquée sur la recherche en biologie de l'évolution, particulièrement sur le développement des approches réseaux en phylogénétique.

Cela étant dit, les réseaux explicites plus récents ne peuvent être compris que partiellement sur la base d'une séparation entre les modèles inférentiels et de représentation. En effet, bien que des réseaux explicites soient encore aujourd'hui utilisés pour représenter des résultats obtenus exclusivement sur la base d'inférences arborescentes comme c'est le cas dans l'article de Doolittle (par exemple Vanderpool et al. 2020), des méthodes inférentielles mobilisant les réseaux sont aussi utilisées activement pour inférer des hypothèses phylogénétiques réticulées. Le programme le plus complet et le plus utilisé (son utilisation demeure faible si on le compare à des programmes équivalents centrés sur le modèle arborescent) pour l'inférence de réseaux phylogénétiques explicites est PhyloNet (Than, Ruths et Nakhleh 2008; Wen et al. 2018). En date du 27 septembre 2021, GoogleScholar recense 343 citations pour la publication originale de PhyloNet (Than, Ruths et Nakhleh 2008) et 138 pour une mise à jour des multiples modifications et des ajouts caractérisant de nos jours le programme (Wen et al. 2018). À titre de comparaison, le programme PhyloNetworks (Solís-Lemus, Bastide et Ané 2017) en compte 116 alors que BEAST2 (Zhang et al. 2018) en compte 98. Il est à noter que ces différents logiciels contiennent des fonctionnalités différentes et mettent de l'avant des approches mathématiques

et statistiques différentes. Dans ce qui suit, je me concentre sur PhyloNet (Wen et al. 2018) afin d'illustrer les grandes lignes de l'inférence de réseaux phylogénétiques explicites sur la base de méthodes réseaux.

PhyloNet est un progiciel développé afin de faciliter l'étude de phénomènes évolutifs produisant de la réticulation au sein de l'histoire évolutive. Plus spécifiquement, cette suite de logiciels permet de produire des représentations réticulées, de caractériser des réseaux phylogénétiques (sur la base de trois métriques distinctes), de comparer des topologies réticulées et d'inférer des réseaux phylogénétiques (Than, Ruths et Nakhleh 2008). Parmi ces quatre fonctionnalités, celle qui m'intéresse tout particulièrement est la dernière (je reviens aussi, plus loin, sur la seconde).

Initialement, une seule méthode de reconstruction de réseaux phylogénétiques était implantée dans le progiciel. Cette méthode, RIATA-HGT (Nakhleh, Ruths et Wang 2005), permet d'inférer les liens phylogénétiques entre lignées qui sont tributaires de transferts génétiques horizontaux. Les données utilisées pour réaliser de telles inférences sont des arbres phylogénétiques préalablement construits : un arbre d'espèces et des arbres de gènes qui ne concordent pas avec l'arbre d'espèce concerné. Le problème consistant à réconcilier ces arbres incompatibles par l'ajout, sur l'arbre des espèces, de liens horizontaux entre les lignées tout en minimisant leur présence est cependant un problème impossible à calculer en une durée raisonnable (ce qui témoigne des limites de la puissance calculatoire disponible) : « The problem of finding a minimum-cardinality set of HGT events whose occurrence on species tree ST would give rise to the gene trees in set T is computationally hard » (Than, Ruths et Nakhleh 2008, 321-322).

Ainsi, une heuristique a été développée (Nakhleh, Ruths et Wang 2005) pour résoudre le problème en travaillant avec l'arbre des espèces et un seul arbre de gènes à la fois. Par exemple, si trois arbres génétiques se révèlent incompatibles avec l'arbre des espèces, une solution basée sur RIATA-HGT est calculée pour chacun de ces trois arbres, puis celles-ci sont combinées : « Our implementation of RIATA-HGT allows the user to specify a set of gene trees, and it iterates over each pair of the species tree and a gene tree, and summarizes the results of all trees. » (Nakhleh,

Ruths et Wang 2005, 88). Contrairement aux réseaux abstraits, les réseaux qui résultent de ce processus contraignent la représentation des incompatibilités par l'entremise d'un critère théorique. Autrement dit, puisqu'il s'agit de réseaux explicites et donc d'hypothèses phylogénétiques concrètes, le champ des possibles gagne à être restreint sur la base de critères théoriques biologiquement justifiés. En l'occurrence, on observe la prévalence du critère de parcimonie selon lequel il est judicieux de présumer que la solution présentant le moins d'arêtes horizontales est celle qui permet de rendre compte des données le plus adéquatement. Ainsi, Than, Ruths et Nakhleh nous disent la chose suivante : « the minimization criterion is a reflection of Occam's razor: in the absence of any additional biological knowledge, HGT events should be used sparingly to explain data features otherwise explainable under a tree model. » (Than, Ruths et Nakhleh 2008, 321). S'il s'agissait de réseaux abstraits visant simplement à représenter les données, il serait difficile de justifier l'utilisation d'une telle contrainte théorique (une instantiation biologiquement justifiée du rasoir d'Ockham). Il est donc important, pour comprendre l'utilisation suggérée de cette méthode, de saisir que les arêtes latérales présentes au sein de ces réseaux explicites sont considérées comme des représentations de potentiels événements biologiques concrets, soit des transferts génétiques entre lignées, des événements que les auteurs croient rares.

Depuis le lancement initial de PhyloNet, d'autres méthodes d'inférence des réseaux phylogénétiques explicites ont été développées et implantées dans le progiciel (Wen et al. 2018), dont trois sont présentées en détail dans l'article de 2018. La première consiste à inférer des réseaux sur la base d'arbres d'espèces et de gènes incompatibles (comme la méthode présente dans le progiciel depuis 2008 le faisait déjà), mais cette fois de sorte à identifier des arêtes horizontales associées à l'hybridation des espèces (plutôt qu'aux transferts génétiques horizontaux). Le tout requiert une adaptation du principe de parcimonie qui favorise la minimisation du nombre de lignées qui sont ajoutées par la réconciliation d'arbres phylogénétiques incompatibles. Le critère d'évaluation mobilisé pour minimiser le nombre de lignées est appelé la minimisation des coalescences profondes (*minimizing the number of deep coalescence criterion* [MDC criterion]; Yu, Barnett et Nakhleh 2013). La seconde méthode

(détaillée dans Yu et al. 2014) travaille aussi avec des arbres phylogénétiques préalablement inférés, mais en mobilisant cette fois les outils statistiques liés au maximum de vraisemblance.

Or, en phylogénétique comme ailleurs, les inférences basées sur le maximum de vraisemblance portent flanc à certaines critiques (Attias 1999). Deux sont d'un intérêt particulier pour notre propos. Premièrement, la vraisemblance est toujours plus élevée pour les topologies plus complexes. Par exemple, les topologies contenant le plus de réticulation rendent inmanquablement mieux compte des données selon ce critère. Cette situation est adéquate lorsque la structure du graphe recherchée est connue d'avance, par exemple lorsque la structure bifurquée du modèle arborescent est acceptée de façon apriorique. Au contraire, lorsque la structure est inconnue ou que l'on cherche à comparer des modèles ayant des niveaux de complexité différents (comme les arbres et les réseaux), ce critère devient inutile (Attias 1999; Wen et al. 2018); je reviens sur ce point en 7.3.3. Deuxièmement, la vraisemblance est calculable seulement pour un sous-ensemble restreint de modèles et de topologies puisqu'elle demande une puissance calculatoire très élevée (Wen et al. 2018).

Parmi d'autres constats détaillés dans le texte (mais moins directement liés à mon propos), les deux constats susmentionnés ont motivé Wen et al. à développer un algorithme permettant l'inférence bayésienne de réseaux phylogénétiques (Wen, Yu et Nakhleh 2017), lui aussi implanté dans PhyloNet. Cette troisième méthode d'inférence est la seule, parmi celles présentées en détail par Wen et al. (2018), à offrir la possibilité d'utiliser comme intrants des données génétiques (des alignements de plusieurs gènes) et non des arbres phylogénétiques. Ce modèle a été développé afin de désavantager la complexité des topologies (pour renverser la tendance des méthodes associées au maximum de vraisemblance à privilégier les topologies les plus réticulées et ainsi à minimiser les chances de faux positifs dans l'identification de réticulation au sein des phylogénies reconstruites), notamment pour éviter que le signal phylogénétique des phénomènes de triage incomplet des lignées (*incomplete lineage sorting*; phénomène qui engendre des signaux confondants dans les inférences phylogénétiques parce qu'un gène est présent sous différentes formes au sein de la population ancestrale) entraîne une prolifération non justifiée de la réticulation. Ainsi, cette troisième méthode répond à des problèmes associés

au maximum de vraisemblance, mais elle permet aussi une transformation notable des itérations épistémiques de la discipline en escamotant toute référence aux arbres phylogénétiques.

En résumé, il y a donc deux défis principaux auxquels fait face le développement de la phylogénétique centrée sur les modèles réseaux. Premièrement, la puissance calculatoire nécessaire pour inférer de telles topologies est très élevée, même lorsqu'un seul type de phénomène générant de la réticulation est pris en compte. En effet, les méthodes présentées précédemment ciblent indépendamment des phénomènes comme l'hybridation ou les transferts génétiques latéraux et sont ainsi incapables d'offrir un regard exhaustif sur la réticulation présente au sein de l'histoire évolutive (Morrison 2014a; Ottenburghs 2019; Wen et al. 2018). Deuxièmement, la plus grande complexité des modèles réseaux engendre *de facto* une correspondance accrue avec les données sans que cela implique la supériorité épistémique de ces modèles. Au-delà du fait qu'elle rend difficile de trancher lesquels des modèles réseaux ou arborescents devraient être préconisés à l'avenir, cette situation explique la potentielle surévaluation de la réticulation qui pourrait être causée par l'utilisation de modèles réseaux.

Maintenant que les principaux outils de l'inférence phylogénétique des réseaux ont été présentés, il serait possible de faire une première comparaison entre les modèles réseaux et le modèle arborescent. Cependant, une telle approche favoriserait la perspective surplombante puisque nous avons pour l'instant seulement appréhendé les modèles réseaux sur la base de leur propre logique. Dans la prochaine section, je resitue ces méthodes au sein de leur trajectoire historique afin de pouvoir réaliser une analyse épistémologique adoptant une perspective synchronique sur les modèles phylogénétiques.

7.2.2 Une brève histoire des réseaux, en phylogénétique et ailleurs

Si une date précise devait être donnée pour la naissance de la théorie des graphes, et plus largement de l'étude mathématique des réseaux, ce serait 1735. Le mathématicien Leonard Euler (1707-1783) offrait alors une résolution au problème des ponts de Königsberg. La ville de Königsberg, ancienne capitale prussienne aujourd'hui appelée Kaliningrad, s'étend sur quatre étendues de terre : deux îles formées par les bras de la Pregolia et les deux rives de ce fleuve. Ces quatre étendues étaient à l'époque liées par sept ponts. Le problème mathématique ayant donné

naissance à la théorie des graphes consistait à déterminer s'il est possible de traverser chacun des sept ponts sans passer deux fois sur le même. Dans un langage contemporain de la théorie des graphes, il s'agit de savoir si un chemin peut passer par chacune des arêtes sans repasser deux fois par la même; les arêtes sont les ponts et les nœuds sont les étendues de terre. Euler a rigoureusement démontré, en 1735, l'impossibilité qu'un tel chemin existe. La théorie des graphes est née à ce moment, bien qu'elle n'ait pris son envol qu'au 20^e siècle (Barabási 2016; Newman 2018).

À la même époque, mais indépendamment des développements mathématiques en question, les pratiques classificatoires des naturalistes étaient dominées par des représentations du monde biologique truffées de réticulation (Morrison 2014b; Ragan 2009; Stevens 1984, 1994). Bien qu'il soit tout de même erroné d'avancer que le modèle réseau soit plus vieux que le modèle arborescent (« The network metaphor thus [...] pre-dates the tree metaphor » [Morrison 2014b, 629]), ce dernier datant de l'Antiquité (Crombie 1994; Gontier 2011; Panchen 1992), il est très certainement adéquat d'affirmer que les deux modèles précèdent historiquement la mise en place du cadre darwinien. De surcroît, il est important de souligner que l'utilisation de représentations réticulées, au 18^e siècle tout comme aujourd'hui, s'ancre dans une logique d'exhaustivité (Tort 1989) : les classifications hiérarchiques ou linéaires (la *scala naturae*) brossent un portrait limité des relations entre individus ou groupes taxonomiques; les réseaux permettent de mieux représenter la grande complexité des relations qui animent le monde vivant. Les travaux d'Adanson et le rôle de ce naturaliste dans la mise en place de la « méthode naturelle » offrent une fenêtre privilégiée pour comprendre cette logique de l'exhaustivité. Mes propos, dans ce qui suit, sont fortement inspirés de l'étude des travaux d'Adanson réalisée par Tort (1989).

Adanson (1757, 1763) était critique des systèmes de classification artificiels, comme celui de Linné qui jouissait à l'époque d'un soutien grandissant. L'artificialité de tels systèmes classificatoires découle de la sélection de traits sur la base desquels la classification devrait être organisée, une sélection nécessaire à la mise en place d'une hiérarchie de catégories emboîtées les unes dans les autres (Rieppel 2010a; Stevens 1997). Linné, par exemple, priorisait les traits relatifs à la reproduction. En se limitant à un nombre limité de traits, il est plus probable que les

affinités entre organismes inférés sur la base de l'analyse de ces traits soient cohérentes. Si, au contraire, les affinités inférées sont incohérentes les unes avec les autres (ce qui est plus probable lorsqu'on considère une plus grande variété de traits), alors il devient impossible de ranger les organismes dans une hiérarchie bien ordonnée qui traduirait ces affinités. Au 18^e siècle comme aujourd'hui, la classification requiert un tri apriorique (théoriquement justifié) des critères retenus pour son opérationnalisation.

Adanson, en développant la méthode naturelle, cherchait à rejeter la partialité des classifications artificielles de son temps, à éviter la simplification des analyses de similarité tout juste décrite :

La 'méthode naturelle' introduit en botanique le rêve d'une classification sans figure, purement objective et non partielle, échappant au particularisme du point de vue, car soumise à la catégorie de la *totalité* : *toutes* les parties doivent être soumises à l'examen, et la démarche classificatoire se subordonne au projet d'*exhaustivité descriptive et comparative*. (Tort 1989, 209, italiques présents dans le texte original)

En ce sens, la méthode descriptive mise de l'avant par Adanson est une actualisation de ce que j'ai décrit plus tôt dans cette thèse (chapitre 3) par l'expression *naturalisme*, soit l'idée lockéenne selon laquelle les catégories scientifiques devraient se conformer à la nature telle qu'elle existe indépendamment de la subjectivité des chercheurs et des chercheuses. Cette injonction de naturalité menait, chez Adanson, à une injonction d'exhaustivité. Selon Tort, Adanson était conscient de l'impossibilité d'atteindre une exhaustivité parfaite, une comparaison des organismes basée sur l'ensemble de leurs êtres, mais insistait sur l'importance de cette exhaustivité en tant qu'idéal à atteindre. Concrètement, cela impliquait le rejet des classifications explicitement et volontairement partielles au profit de classifications plus étoffées.

Une autre conséquence du désir d'exhaustivité associé au développement de la méthode naturelle a été la popularisation des représentations réticulées. En effet, l'analyse des organismes sur la base d'une pluralité de traits met inmanquablement en lumière des affinités pluridirectionnelles entre organismes, des affinités incompatibles avec le modèle hiérarchique. Si Adanson espérait réconcilier exhaustivité et naturalité avec classification hiérarchique, les naturalistes de son époque et même ceux et celles lui ayant succédé n'hésitaient aucunement à aplanir les systèmes d'affinités entre espèces afin d'illustrer toute leur complexité. Alors qu'il

traite de cette tendance des 18^e et 19^e siècles à représenter les relations entre espèces de façon réticulée (et du rapport de cette tendance à la *scala naturae*), Stevens nous dit la chose suivante :

The idea that relationships were simply linear broke down as more numerous characters, especially those of anatomy, were used in classification, and the language of highness and lowness would seem to be inappropriate for this new world in which relationships were at the very least reticulating. (Stevens 1994, 164-165)

Morrison abonde en ce sens :

The network (or web or map) was the predominant metaphor during the 1800s » (Morrison 2014b, 629). Même Linné, lorsqu'il cherchait à représenter les relations naturelles (en utilisant la métaphore de la carte géographique) plutôt qu'à construire un système classificatoire utile et pratique, soulignait le fait que les liens entre espèces sont pluridirectionnels : « All plants show affinities on either side, like territories in a geographical map. (Linné 1751, cité dans Morrison 2014b, 629)

Ainsi, les représentations réticulées du monde vivant ont joué un rôle important en systématique jusqu'à la mise en place de la pensée arborescente. Comme présenté au chapitre précédent, la deuxième moitié du 19^e siècle et surtout le 20^e siècle ont vu le modèle arborescent et la classification hiérarchique qui l'accompagne être projetés au cœur de la systématique biologique, notamment sous l'influence de la pensée darwinienne puis néo-darwinienne. Il a fallu attendre les années 1970 (Grant 1971; Sneath 1975) et plus significativement le tournant du 21^e siècle pour que les représentations réticulées, à travers les modèles réseaux, retrouvent une place légitime en systématique.

En effet, les succès associés à la transition darwinienne avaient en leur cœur la notion de divergence des lignées, comme mentionné préalablement dans cette thèse. Ce qui est intéressant de noter est que la logique invitant la réintroduction de la réticulation est, d'un point de vue abstrait, la même que celle ayant justifié l'adhésion des naturalistes à ce mode de représentation aux 18^e et 19^e siècles : les modèles hiérarchique et arborescent sont incapables de rendre compte des observations empiriques de façon satisfaisante. À l'époque, il s'agissait simplement de tenir compte du plus grand nombre de données possible, de se soumettre à l'injonction d'exhaustivité (dans une logique qui rappelle la preuve intégrale, mais qui s'actualise par des représentations réticulées); de nos jours, il s'agit de rendre compte de phénomènes évolutifs bien précis, soit ceux

qui s'opposent à la bifurcation des lignées et qui sont incompatibles avec la simplification des phénomènes sous-jacente à la pensée arborescente.

Plus spécifiquement, les modèles réseaux ont d'abord été mobilisés en tant qu'outils mathématiques permettant de manipuler et de représenter des données incohérentes. Pour rappel, des données morphologiques ou génétiques sont dites incohérentes, dans une perspective traditionnelle, si elles soutiennent des topologies arborescentes distinctes. Ainsi, les réseaux de bifurcation mentionnés dans la section précédente ont été utilisés en phylogénétique jusque dans les années 1990 sans être considérés comme des outils spécifiquement phylogénétiques :

it was Bandelt [1994] who formalized the reference to unrooted networks as 'phylogenetic networks'; previously, these had been simply called 'split graphs'. This use of the term 'phylogenetic' does not refer to genealogies, since the graphs are unrooted and thus have no time dimension. (Morrison 2014a, 1523)

Ce n'est que plus tard que ces réseaux en sont venus à être considérés comme phylogénétiques à proprement parler, alors qu'émergeait l'idée selon laquelle des phénomènes biologiques pertinents pour comprendre l'évolution pouvaient bel et bien expliquer ces soi-disant incohérences entre données.

Parmi les avancées ayant imposé la notion de réseau phylogénétique, deux ressortent du lot. Premièrement, la suggestion, maintenant largement acceptée, selon laquelle la relation entre les trois grands domaines du monde vivant (eucaryotes, archées et bactéries) serait elle-même réticulée. Les eucaryotes seraient apparus à la suite d'un phénomène d'endosymbiose ayant lié le destin évolutif d'une lignée d'archées et d'une de bactéries plutôt que par un phénomène de divergence (Doolittle 2000). L'observation d'autres endosymbioses (comme celle menant à la photosynthèse eucaryote) est venue renforcer ce constat. Deuxièmement, la réalisation de l'omniprésence des phénomènes de transmission génétique horizontaux (ou latéraux), au moins au sein des archées et des bactéries, sans parler du monde viral, a aussi été cruciale dans la justification de l'utilisation de modèles réseaux (Baptiste et al. 2004; Doolittle 1999; Velasco 2012). Bien que cela ne concerne qu'une fraction de l'utilisation des réseaux en phylogénétique, certains de ces réseaux sont ainsi devenus plus que des outils mathématiques : la prise en compte

de phénomènes évolutifs considérés comme étant théoriquement pertinents permet de transformer une information préalablement considérée comme du bruit en information qui mérite d'être explicitement représentée au sein des modèles phylogénétiques. En ce sens, une partie de l'utilisation des réseaux d'aujourd'hui s'inscrit en continuité avec la méthode naturelle d'Adanson. Les réseaux explicites sont considérés comme des représentations plus exhaustives que les autres modèles disponibles.

7.2.3 Une première analyse comparative des itérations contemporaines de la phylogénétique : les approches réseaux et la pensée arborescente

L'utilisation des modèles réseaux pour appréhender des quantités toujours grandissantes de données phylogénétiques représente l'itération la plus récente de la systématique. Il convient maintenant de la comparer avec l'itération précédente (les deux se chevauchant actuellement), soit l'approche phylogénétique (ou phylogénomique, si on veut insister sur la quantité de données traitées) associée à la pensée arborescente, comme présentée au chapitre 6. Le portrait que je brosse de ces deux itérations est le fruit d'analyses discursives qui m'ont permis de reconstruire des postures générales. Il va sans dire que les constructions théoriques qui en résultent ne correspondent pas exactement à chaque itération locale mobilisant réseaux ou arbres. Les résultats de mon étude peuvent tout de même servir de grille d'analyse pour appréhender des pratiques phylogénétiques contemporaines, à condition de l'adapter aux spécificités des textes ainsi approchés.

Dans ce qui suit, j'amorce l'analyse comparative en traçant les démarcations majeures entre les deux itérations ainsi qu'en soulignant les nombreux points de contact qui se cachent sous l'opposition forte entre les deux modèles. Le tableau 7.1 synthétise l'analyse comparative effectuée; la liste des enjeux comparés est une liste ouverte et flexible que j'encourage à adapter à l'étude de pratiques phylogénétiques subséquentes. Dans la deuxième partie du chapitre, j'explore certains des principaux débats philosophiques relatifs à ces deux modèles.

L'utilisation de modèles arborescents en phylogénétique consiste rarement, voire jamais, à nier que l'histoire évolutive est au moins partiellement réticulée. DeSalle et Riley, par exemple, reconnaissent la présence de transferts génétiques horizontaux au sein des lignées bactériennes,

mais affirment que cela ne nous empêche aucunement de penser leur évolution sur la base du modèle arborescent (l'acronyme HGT réfère, dans la citation suivante, aux transferts génétiques horizontaux) : « if HGT either exists at a lower level than is thought or can be shown to be surmountable in the context of phylogenetic analysis, then there is no reason to 'uproot' the tree of life. » (DeSalle et Riley 2020, 1178-1179). Comme l'illustre cette citation, le modèle arborescent conserve son rôle central au sein de la discipline si la réticulation (ici réduite aux transferts génétiques horizontaux) est suffisamment anecdotique pour qu'il soit possible de la surmonter pour inférer des topologies arborescentes.

Nonobstant la reconnaissance partagée de l'existence de phénomènes menant à la réticulation phylogénétique, les deux itérations sont irréconciliables sur certains points, notamment sur le plan des représentations des liens phylogénétiques. Les phénomènes menant à la réticulation sont évidemment pluriels (hybridation des espèces, recombinaison génétique, transferts génétiques latéraux, [endo]symbiose, etc.), mais tous peuvent être représentés sur un graphe phylogénétique (dirigé) si on permet qu'un nœud ait un degré entrant d'au moins deux (qu'au moins deux arêtes mènent à ce nœud, et donc que le groupe taxonomique ainsi représenté ait au moins deux ancêtres). Le modèle arborescent, au contraire, ne permet que des degrés entrants d'un ou moins (la racine ayant un degré entrant de zéro). Il y a donc un traitement drastiquement différent des phénomènes menant à la réticulation que je traduis ici par son inclusion, ou non, au sein des représentations phylogénétiques. Cette inclusion témoigne d'un changement théorique important relatif à la primauté de la divergence au sein de la pensée évolutionniste : selon la phylogénétique centrée sur les modèles réseaux, réduire l'évolution biologique aux phénomènes de divergence nuit à la bonne compréhension des phénomènes étudiés. La phylogénétique traditionnelle, au contraire, considère qu'il faut « surmonter » la réticulation pour extraire un patron divergent à partir d'un ensemble de données.

Tableau 7.1 – Synthèse de l’analyse comparative de la phylogénétique traditionnelle et de la phylogénétique centrée sur les réseaux

	Enjeux	Phylogénétique traditionnelle	Phylogénétique centrée sur les modèles réseaux
Forte opposition	Représentation de l’histoire évolutive	Modèle arborescent	Modèle réticulé
	Importance accordée au principe de divergence	Fondamentale	L’importance du principe de divergence est nuancée par l’acceptation de la réticulation
	Traitement des données dites conflictuelles	Elles sont considérées comme du bruit	Il s’agit, dans certains cas, d’information permettant d’inférer l’existence d’événements évolutifs générant de la réticulation
Opposition nuancée	Importance inférentielle de la réticulation	Nulle	Limitée (les degrés entrants sont habituellement limités à 2)
	Importance inférentielle du modèle arborescent	Le modèle structure les inférences	La majorité des méthodes utilisent des arbres de gènes et d’espèces comme intrants
	Principe d’histoire évolutive unique	Présent lorsque les arbres enracinés/explicites sont utilisés, absent lorsqu’on se limite à des arbres non enracinés/abstraites; la tendance au réalisme envers l’Arbre du vivant est bien présente et forte	Idem; la tendance au réalisme est plus nuancée
Aucune opposition	Structure inférentielle (à un haut niveau d’abstraction)	La similarité est utilisée pour inférer la phylogénie	Idem
	Analyse de la similarité	Avec ou sans alignement	Idem
	Importance des lignées pour comprendre l’évolution	Fondamentale	Idem
	Méthode descriptive priorisée	Classification (négligeant les différences entre individus)	Idem

Le tableau 7.1 synthétise les points saillants de l’analyse comparative des itérations phylogénétiques centrées sur les arbres et les réseaux. En bleu, j’ai identifié les enjeux sur lesquels les deux approches se rejoignent complètement. En gris, j’ai identifié les enjeux plus nuancés alors que la couleur orange est utilisée pour souligner les différences majeures entre les deux itérations.

D'un point de vue inférentiel, le rapport entre les deux modèles est plus nuancé. Considérons le relâchement des contraintes imposées aux degrés entrants des nœuds des graphes phylogénétiques tout juste mentionné. Ce relâchement, autant présent sur le plan inférentiel que celui des représentations de l'histoire évolutive, est partiel. Que ce soit pour des raisons calculatoires (Huson, Rupp et Scornavacca 2010) ou théoriques (afin de conserver l'idée de lignée intacte; j'y reviens plus loin dans cette section et en 7.3.5), le degré entrant est habituellement limité à deux : « it is generally assumed that only two species can be involved in any such event and so sometimes the set of possible indegrees of nodes is restricted to the values 0, 1 and 2, corresponding to the root, tree nodes and reticulations involving only two in-edges, respectively » (Huson, Rupp et Scornavacca 2010, 139-140). Un groupe taxonomique peut avoir un maximum de deux ancêtres directs.

Cette contrainte, tout à fait compréhensible d'un point de vue pratique (question de garder le champ des possibles sous contrôle) peine à s'imprégner dans l'esprit d'une personne s'étant familiarisée avec l'étendue des transferts génétiques horizontaux (Baptiste et al. 2004; Baptiste et Boucher 2008; Doolittle 1999; Keeling et Palmer 2008; McDaniel et al. 2010) ou des phénomènes de recombinaison génétique (Méheust, Lopez et Baptiste 2015; Méheust et al. 2018; Naor et al. 2012; Toleman et al. 2012). La limitation du nombre d'ancêtres à deux pourrait être une contrainte calculatoire avec peu de fondements biologiques. Les avantages calculatoires d'une telle contrainte sont toutefois significatifs puisque cette limitation rend possible l'inférence phylogénétique basée sur les réseaux (explicites). Cette analyse me pousse à réitérer que les données et la dimension théorique sont insuffisantes pour comprendre ce qui dicte la pratique. Ici, le besoin de rendre observable ce qui ne l'est pas (c'est-à-dire, dans le cas qui nous intéresse, de calculer la topologie phylogénétique sur la base de liens de similarité) est l'un des déterminants importants du modèle. Autrement dit, un modèle impossible à opérationnaliser serait difficile à défendre.

Sur le plan inférentiel, d'autres éléments de comparaison sont à mentionner. En effet, si certaines méthodes d'inférence passent directement des données génétiques aux réseaux phylogénétiques (Wen et al. 2018), la majorité des modèles réseaux sont inférés sur la base d'arbres phylogénétiques (Huson, Rupp et Scornavacca 2010; Wen et al. 2018). L'ensemble des

outils d'analyse de la similarité et de reconstruction d'arbres est donc récupéré par l'itération phylogénétique centrée sur la réticulation. La différence majeure entre ces pratiques et les précédentes relève d'un ajout, d'une étape supplémentaire qui est celle visant à représenter les conflits entre données (comme avec les réseaux de bifurcation) ou à inférer des événements phylogénétiques (hybridation, transferts génétiques latéraux, etc.) qui expliquent ces conflits entre données (qui ne sont dès lors plus des conflits).

Il faut donc souligner une certaine continuité entre les deux itérations étudiées. Plus précisément, l'itération associée aux modèles réseaux vient remplir des lacunes de l'itération précédente (principalement en traitant de l'information que les modèles arborescents peinent à gérer) *tout en s'inscrivant en continuité avec elle*. Là où la phylogénétique plus traditionnelle, guidée par la croyance moniste sous-jacente, exclut nécessairement une partie des données conflictuelles afin de représenter une histoire évolutive unique et arborescente, les modèles réseaux offrent une avenue pour intégrer une partie de ces données conflictuelles *tout en satisfaisant l'injonction du principe d'histoire évolutive unique* (du moins lorsque l'objectif est de générer des réseaux explicites). Il s'agit ici d'une nette illustration du besoin constant de négocier entre l'injonction au progrès (ouverture du champ des possibles en relâchant la contrainte de la bifurcation), qui invite la rupture avec le passé, et le principe de respect, qui appelle la stabilité (analyses de similarité mobilisant les mêmes méthodes, inférences mobilisant les arbres pour obtenir des réseaux, principe d'histoire évolutive unique, etc.). Cette négociation est d'autant plus ardue que, comme le problème de la coordination le résume bien, la validité empirique des modèles phylogénétiques impliqués dans cette dynamique de progrès ne peut être établie que par leur utilisation (j'y reviens en détail en 7.3.3).

En mentionnant que l'utilisation des modèles réseaux se conforme au principe d'histoire évolutive unique, j'ai souligné une ressemblance entre les deux itérations phylogénétiques à l'étude. Celle-ci doit cependant être nuancée puisque les auteurs et autrices associé-e-s aux modèles réseaux tendent à considérer le rapport entre ces modèles et l'histoire évolutive « réelle » de façon plus ambivalente. Par exemple, Than et al. insistent sur le fait que les réseaux sont d'abord et avant tout des modèles qui permettent de rendre compte des différents phénomènes biologiques ayant un impact sur l'histoire évolutive : « However, when *reticulate*

evolutionary events such as horizontal gene transfer or interspecific recombination occur, the evolutionary history is more appropriately modeled by an evolutionary network. » (Than, Ruths et Nakhleh 2008, §1, italiques présents dans l'original). Pourtant, le paragraphe suivant du même texte se veut plus explicite quant à la teneur réaliste de leur approche centrée sur les réseaux, en insistant sur le fait que les relations phylogénétiques sont bel et bien réticulées et donc que la réticulation est plus qu'un outil épistémique : « All of these processes result in networks, rather than trees, of evolutionary relationships » (Than, Ruths et Nakhleh 2008, §2). Finalement, dix ans plus tard, le groupe de chercheurs et de chercheuses mené par Nakhleh se montre explicite quant à la posture réaliste qu'ils et elles adoptent à l'égard des réseaux phylogénétiques : « Needless to say, one is interested in the true graph structure and not one that is more complicated simply because it assigns higher likelihood to the data. » (Wen et al. 2018, 739). Les méthodes statistiques nous permettant d'inférer les relations phylogénétiques sous la forme de réseaux nous offriraient, selon cette posture, un regard sur la « vraie structure des graphes [phylogénétiques] ». Nul besoin de préciser, à ce point-ci de la thèse, que cette posture est incompatible avec les conséquences du problème de la coordination.

Nonobstant cette ambivalence, il convient de souligner que les modèles réseaux sont régulièrement considérés comme des modèles, des représentations abstraites du phénomène (les relations phylogénétiques) qui favorisent son explication, qui sont réalisés sur la base de diverses contraintes théoriques. Si c'est aussi le cas pour les arbres (et notamment pour les arbres non enracinés), la tendance semble plus forte pour les réseaux (peut-être en raison de leur importation récente en biologie, contrairement à l'arbre qui siège au cœur de la discipline depuis suffisamment longtemps pour y avoir été sédimenté, ontologisé, réifié).

Nous pouvons continuer l'analyse comparative en observant la structure inférentielle des deux itérations de la systématique à un haut niveau d'abstraction. Cette structure est la même : dans les deux cas, il s'agit d'inférer des relations phylogénétiques entre unités taxonomiques, et ces inférences sont réalisées sur la base d'analyses de similarité. Ainsi, un nombre important de forces et de faiblesses sont partagées par ces deux approches. L'attraction des longues branches, par exemple, est un artéfact auquel sont vulnérables les approches réseaux tout autant que les

approches traditionnelles puisqu'il s'agit d'un problème pour les méthodes d'alignement multiple.

Dans le même ordre d'idées, un fait évident mais crucial doit être souligné. Les deux approches considèrent que l'étude de l'histoire des lignées biologiques est une porte d'entrée adéquate pour l'étude de l'évolution (et d'autres phénomènes biologiques). La discipline, dans son ensemble, vise à reconstruire l'histoire du monde vivant, une histoire cependant réduite à celle de lignées de gènes ou d'organismes, une histoire qui fait abstraction, entre autres choses, des interactions entre diverses entités biologiques à différents niveaux d'organisation ainsi que de l'histoire de l'environnement. Je reviens sur ce point en détail en 7.3.4.

Un dernier élément de comparaison, tout aussi évident et important que le précédent, gagne à être mentionné. Les différentes itérations de la systématique décrivent le monde biologique par l'entremise de la classification, c'est-à-dire en découpant et en discriminant le monde biologique, puis en mettant en relation les catégories ainsi obtenues. Tout au cours de cette thèse, j'ai accordé une importance marquée à la création de systèmes. Or, les descriptions axées sur la classification partagent une autre caractéristique, soit celle de décrire les occurrences d'une catégorie sur la base de leur appartenance à cette catégorie. Ainsi, l'ensemble des membres associés à une unité taxonomique sont représentés par un (ou parfois quelques) membre(s), une occurrence censée fournir l'information nécessaire à la compréhension de l'ensemble de la population. Ainsi, une double synecdoque est mobilisée : les unités taxonomiques sont représentées par une seule entité et les liens de similarité sont établis entre espèces sur la base d'un éventail réduit de traits (génétiques ou morphologiques, souvent les deux). De cette double synecdoque, de ces descriptions qui traitent une partie (des parties d'organismes et des organismes) comme étant le tout (l'organisme dans son ensemble et l'unité taxonomique dans son ensemble), une simplicité qui peut certainement favoriser les inférences et la compréhension du monde vivant émane. Il est toutefois indéniable qu'en œuvrant de la sorte, en décrivant le monde biologique sur la base des pratiques classificatoires, une perte d'information s'ensuit. En effet, les différences entre les entités sont négligées du fait même de leur description sur la base de leur catégorie d'appartenance, c'est-à-dire en focalisant sur les points communs entre elles; une perte d'information résulte aussi nécessairement de l'exclusion de nombreux traits pour

l'analyse de similarité sur la base de contraintes théoriques. Les modèles réseaux, tout comme les modèles arborescents, ont donc en commun de s'abstraire de la complexité inhérente des phénomènes à l'étude en décrivant les êtres biologiques sur la base de leur catégorie d'appartenance plutôt que sur leurs particularités. Je reviens en détail sur cette question en 7.3.5.

En résumé, les différences principales entre les deux itérations de la systématique étudiées portent sur le traitement des données qui suggèrent que l'histoire évolutive serait réticulée. Selon les modèles réseaux, ces données peuvent être interprétées efficacement comme étant plus que du bruit; il en résulte que la réticulation devrait être présente au sein des représentations de l'histoire du monde vivant. En contraste, la phylogénétique traditionnelle maintient la primauté de la divergence. Ces différences témoignent d'une ouverture logique qui consiste à accepter l'influence d'une plus grande variété de processus sur l'histoire évolutive et donc à ouvrir certaines boîtes noires de la biologie de l'évolution. Cette ouverture logique a été amplifiée par un corpus de textes en biologie de l'évolution qui mobilise les outils issus de la théorie des graphes afin d'appréhender l'évolution de façon plus inclusive en explorant au-delà de la formation de lignées verticales ou quasi verticales (Corel et al. 2016; Doolittle et Inkpen 2018; Bapteste et Huneman 2018; Bapteste et Papale 2021; Papale, Saget et Bapteste 2020; Suárez 2020; Watson et al. 2019). Au sein de la phylogénétique centrée sur les réseaux, l'ouverture logique ne bée toutefois pas aussi grande. Au contraire, j'ai démontré, par l'analyse comparative précédente, que les arbres et les lignées verticales occupent toujours une place centrale dans cette itération. L'opposition forte que le corpus disciplinaire suggère entre modèles réseaux et modèles arborescents se doit d'être nuancée.

7.3 Les débats entourant les réseaux phylogénétiques

Dans cette seconde partie du chapitre, je présente cinq débats philosophiques dont les enjeux sont directement liés aux modèles phylogénétiques arborescents et réseaux. Les trois premiers sont des débats concernant la validité épistémique des réseaux phylogénétiques et la possibilité de choisir le meilleur type de modèles pour assurer le progrès de la discipline; il s'agit de débats déjà présents au sein de la phylogénétique contemporaine ou du corpus de philosophie de la biologie s'intéressant à cette discipline. Les deux autres débats sont absents du corpus en

question, mais concernent des enjeux que je considère comme étant importants pour les pratiques en biologie de l'évolution actuelles et futures. Ces débats concernent des points de rencontre entre les deux types de modèles et les implications philosophiques de ces similarités.

7.3.1 De la compatibilité entre les réseaux phylogénétiques et le cadre darwinien

Les nombreuses données issues de la génomique et le traitement de ces données par l'entremise de modèles réseaux ont suscité de vifs débats en biologie de l'évolution. Une partie de ces débats a été structurée sur la base d'une potentielle remise en question du cadre de pensée darwinien engendrée par la génomique. Selon une partie de la communauté, les nouveaux processus biologiques, notamment les transferts génétiques horizontaux, promettent une révolution qui pourrait dissocier l'étude de l'évolution et le nom de Darwin : « Thus, we regard as regrettable the conventional concatenation of Darwin's name with evolution, because other modalities must also be considered. » (Goldenfeld et Woese 2007, 369). Pour d'autres membres de la communauté, le rejet du darwinisme est prématuré :

Darwin's ideas seem therefore to be alive and well. However, in recent years, whereas still being a devil for religious fanatics, Darwin became the target of heavy criticisms coming from part of the scientific community itself. Several genomists and molecular evolutionists have argued that genomic data have challenged Darwin's view of life (Baptiste et al., 2009; Dagan and Martin, 2009; Doolittle, 2009; Koonin, 2009a, 2009b; Raoult, 2010). Notably, they have suggested that 'Darwinism' is only valid for eukaryotes not for prokaryotes (assimilated to microbes) and proposed to replace the 'Tree of life' (TOL, supposed to be the hallmark of Darwinism, but see Penny, 2011) by networks (or rhizome) to take into account gene flows between organisms (Baptiste et al., 2009; Dagan and Martin, 2009; Raoult, 2010). In a review title, Ford Doolittle wondered what 'the demise of Charles Darwin's tree of life hypothesis means for classification and the theory of evolution' (Doolittle, 2009). Lamarck has been (once more) awoken to confront Darwin, as illustrated for instance by the title of another recent review paper: 'Is evolution Darwinian or/and Lamarckian?' (Koonin et Wolf, 2009). (Forterre 2012, §2)

Une question se pose donc : l'utilisation de modèles réseaux, justifiée par des données issues de la génomique, représente-t-elle un désaveu du cadre de pensée darwinien? Si celui-ci est réduit à la pensée arborescente dans son ensemble, alors il semblerait que oui. Au chapitre précédent, j'ai cependant remis en question la validité d'un tel amalgame. Or, même après une analyse détaillée de la pensée darwinienne, des biologistes considèrent que le renouveau de la

pensée évolutionniste du 21^e siècle se veut effectivement anti-darwinien (Baptiste et al. 2009; Dagan et Martin 2009; Doolittle 2009; Goldenfeld et Woese 2007; Raoult et Koonin 2012; Szathmáry 2015). Prenons comme exemple un texte allant en ce sens publié par Merhej et Raoult en 2012.

Merhej et Raoult (2012) analysent les changements induits par la génomique sur la conception (néo-)darwinienne de l'évolution. Plus spécifiquement, on retrouve dans cet article une série de sept enjeux au sujet desquels la conception classique de l'évolution et le nouveau phylogénomique s'opposent. Premièrement, la conception traditionnelle considère que la généralisation de traits avantageux au sein d'une population, sous l'effet de la sélection naturelle, représente le vecteur premier de l'évolution. Deuxièmement (et conséquemment), cette vision de l'évolution mène à concevoir les entités biologiques comme étant bien adaptées à leur environnement. Troisièmement, cette évolution mènerait à la complexification des entités biologiques, et quatrièmement, cette complexification résulterait de l'accumulation de légères mutations au sein du génome et du phénotype. Cinquièmement, la conception darwinienne de l'évolution est ancrée dans la transmission verticale de traits hérités au sein de lignées parent-rejets. Sixièmement, tous les organismes cellulaires auraient un ancêtre commun. Finalement, l'ensemble des dynamiques tout juste décrites pourraient être représentées par un seul arbre phylogénétique, l'Arbre du Vivant.

Selon Merhej et Raoult, les avancées issues de la génomique (et notamment de la phylogénomique) remettent en question chacun de ces points. Premièrement, l'effet de la sélection naturelle n'est qu'un processus parmi d'autres guidant l'évolution des populations. Deuxièmement, les génomes contiennent une quantité importante de matériel génétique qui ne contribue aucunement à l'adaptation des organismes à leur environnement (et donc que la sélection naturelle ne peut pas expliquer). Troisièmement, les organismes complexes ne représentent qu'une petite partie du monde vivant; de surcroît, nombre de lignées d'organismes pluricellulaires se sont éteintes par le passé, ce qui fait douter du fait que l'évolution mène nécessairement à la complexification. Quatrièmement, la variation sujette à la sélection naturelle est souvent générée de façon marquée et non graduelle (duplications, délétions, insertions ou réarrangements de longs segments de génomes). Cinquièmement, le matériel génétique des

organismes peut être reçu par des mécanismes menant à la réticulation phylogénétique (transferts génétiques horizontaux, par exemple) et non seulement dans une dynamique de création de lignées verticales parents-rejetons. Sixièmement, il est impossible d'identifier un ancêtre commun par l'entremise de l'analyse des gènes connus. Septièmement, le génome de tout être vivant est un ensemble composite de gènes détenant des histoires évolutives distinctes et significatives pour expliquer divers phénomènes; il est donc impossible de représenter l'évolution par l'entremise de l'Arbre du vivant.

Ainsi, seulement trois de ces enjeux concernent *directement* la phylogénétique. Les quatre premiers concernent plutôt les dynamiques des populations, les modalités de l'évolution telles qu'on peut les conceptualiser et les modéliser au sein de regroupements d'organismes isolés. Les trois derniers, au contraire, traitent de dynamiques de macroévolution et sont directement liés au débat entre le modèle arborescent et les modèles réseaux (certains enjeux de microévolution font leur chemin jusqu'aux pratiques phylogénétiques, mais je centre ici mon propos sur les enjeux qui sont incontestablement phylogénétiques). Selon Merhej et Raoult (entre autres; Woese 2002), l'histoire des lignées aurait, depuis le début, été réticulée et parsemée de transferts génétiques latéraux qui requièrent des représentations et des modèles tout aussi réticulés afin d'être appréhendée adéquatement :

None of the seven points laid out in the introduction to this manuscript can be permanently retained, as established by Darwin's theory [...]. This theory cannot be upheld in its entirety. Recent advances from genomics refute the ideas of gradualism, exclusive vertical inheritance, evolution selecting the fittest, a common ancestor and the TOL. [...] A single tree is largely inadequate. We prefer to represent evolution as a family tree or in the form of a rhizome, which corresponds to a more authentic description of our present knowledge than the TOL. (Merhej et Raoult 2012, 113)

Dans la perspective de Merhej et Raoult, les réseaux phylogénétiques traduisent donc un décalage entre la vision darwinienne de l'évolution et la compréhension actuelle des dynamiques de formation de lignées. Le contraste entre cette citation et celle de Forterre offerte précédemment est marqué. D'un côté, le cadre darwinien est considéré comme étant insuffisant et inapte à rendre compte des phénomènes récemment mis à jour. De l'autre, il est au contraire considéré comme étant aussi présent et fort qu'il ne l'a jamais été. Évidemment, une troisième voie est aussi disponible au sein du corpus de textes en question :

Some hold that the new discoveries only add details to the neo-Darwinian view of evolution whereas others maintain that the basic tenets of (neo)Darwinism have been falsified in the Popperian sense (Popper, 1959); yet others strive to find a middle ground by positing that, although the bulk of today's evolutionary biology consists of data and concepts that simply did not exist even 30 years ago, at the core there remains the key Darwinian principle of descent with modification. (Raoult et Koonin 2012, 1)

Cette troisième voix est plus nuancée et fait écho plus adéquatement à l'analyse épistémologique des pratiques phylogénétiques informée par le cohérentisme historique réalisée en 7.2.3. En effet, si les modèles réseaux, symptômes des avancées de la génomique, des phénomènes empiriques concernés, mais aussi de développements matériels et mathématiques, s'inscrivent dans une dynamique de changement des pratiques soumise à l'impératif du progrès, une partie significative des éléments structurants des itérations passées est conservée lorsque cette façon de pratiquer la phylogénétique est mise de l'avant. En ce sens, la transition caractérisée par ces modèles réseaux est loin de consister en un rejet en bloc de ce qui l'a précédé (incluant le cadre néo-darwinien).

En me basant sur l'analyse épistémologique de la transition vers les modèles réseaux et sur des analyses effectuées ailleurs (Papale, Saget et Baptiste 2020), je soutiens donc la troisième voie, pavée notamment par Koonin et ses collègues (Koonin, Makarova et Wolf 2021; O'Malley et Koonin 2011; Raoult et Koonin 2012). La transformation subie par la biologie de l'évolution se veut progressive, de sorte que les idées du passé sont loin d'être reléguées à l'obsolescence complète par la mise en lumière de nouveaux phénomènes et de nouveaux outils pour les analyser. Nonobstant cette stabilité, indéniable lorsqu'on analyse épistémologiquement la phylogénétique basée sur les réseaux, il peut être problématique de faire preuve de conservatisme à l'égard de modèles dont la validité (voir section 7.3.2) et l'utilité (voir section 7.3.3) s'effritent progressivement. La bonne mesure de transformation et de conservatisme doit ainsi être recherchée avec prudence. La saine enquête permettant de l'identifier requiert, entre autres, une analyse des cadres de pensée qui décompose les amalgames et qui évite de penser des cadres théoriques, comme le cadre darwinien, comme étant des monolithes.

Si ce débat, et plus particulièrement les modalités de sa réalisation, est d'une importance philosophique et biologique marquée, il a aussi une portée sociale qui a été soulignée par les acteurs et actrices cité·e·s depuis le début de cette section. Forterre résume le tout comme suit :

This is a sensitive topic, considering the renewal of creationist thinking in fundamentalist religious circles and the wide publicity given to these 'non-orthodox views.' This was best illustrated by the cover of the 'New Scientists' issue published in January 2009 showing a tree of life superimposed with the sentence: 'Darwin was wrong.' Although the existence of anti-Darwinists in the political arena is certainly not a reason to hide fierce debates between evolutionists over mechanisms and representations of evolution, one can regret to see the name of Darwin used as a foil in these debates. After all, nobody said: 'Mendel was wrong' because his concept of the gene was quite different from what we know today. (Forterre 2012, §4)

Depuis sa première itération, notamment sous les impulsions de biologistes comme Haeckel, la théorie de l'évolution par voie de sélection naturelle a été manipulée au sein d'arènes sociopolitiques variées, pour le meilleur et pour le pire. Parmi les discours politiques l'ayant récupérée pour asseoir leur légitimité, on retrouve les discours séculiers visant à redéfinir la place de l'être humain dans l'univers, et du même coup, de la religion dans les sociétés occidentales. Cette tendance est toujours présente aujourd'hui, alors que des biologistes réputé-e-s confrontent cette théorie aux discours religieux (notamment aux discours créationnistes).

Or, s'il y a effectivement un danger, d'un point de vue social, face à la soi-disant montée du créationnisme évoquée par Forterre (il n'offre aucune donnée pour soutenir ce constat), ce n'est certainement pas la remise en question de certains principes darwiniens (le darwinisme vivait, dans la deuxième moitié du 19^e siècle, une crise [voir Gayon 1992] qui ne l'a aucunement empêché de jouer son rôle social, notamment en Allemagne sous l'impulsion de Haeckel). Je suggère que ce qui mine la puissance du discours scientifique, c'est plutôt le fait de mettre sur un pied d'égalité les discours aspirant à la connaissance dite scientifique et les discours à tendance dogmatique en présentant les discours scientifiques comme des discours dogmatiques. L'une des manières d'éviter que les discours scientifiques n'offrent un tel visage est de proscrire les amalgames d'idées. Lorsque l'on présente une grande diversité de principes de biologie de l'évolution comme étant liés au sein d'un monolithe, soit la pensée (néo)darwinienne, on véhicule un tel amalgame. Un tel geste est épistémologiquement non valide; il faut décomposer ces assemblages. Présenter le débat entre réseaux (et données motivant ces réseaux) et arbres (et données justifiant le conservatisme) comme une remise en question de « La pensée darwinienne est inadéquat »; seulement certains principes sont remis en question.

Adopter une posture dogmatique nuisible au progrès scientifique, c'est aussi présenter le discours scientifique comme un ensemble de croyances à considérer comme des vérités absolues. Une telle pratique place les discours scientifiques sur un pied d'égalité avec des discours moins aptes à progresser, d'un point de vue épistémologique. Le progrès scientifique est composé d'une série d'itérations, au sein de laquelle chacune tire une partie de sa légitimité des précédentes (principe de respect) tout en différant au moins légèrement, de sorte à pallier certaines des difficultés identifiées au sein des itérations précédentes (impératif du progrès). Le progrès scientifique dépend donc de la capacité à réviser certaines des croyances constitutives des systèmes épistémiques; autrement dit, les pratiques scientifiques efficaces sont réflexives. De ce fait, il faut être prudent, d'un point de vue universitaire, face à tout discours se proclamant comme une vérité et à toute ambition de faire plus que ce qui est empiriquement possible. En voici un exemple : « Today the theory of evolution is about as much open to doubt as the theory that the earth goes round the sun. » (Dawkins 2006, 30). Le constat de l'existence d'un zèle scientifique non justifié (des éléments de la théorie de l'évolution ont été, sont et devront rester ouverts au doute constructif) m'amène à traiter, dans la prochaine section, d'un autre débat crucial potentiellement constitutif d'un tel zèle. Il s'agit du débat entourant le réalisme à l'égard des modèles phylogénétiques.

7.3.2 Réalisme et modèles phylogénétiques

Comme démontré au chapitre précédent, la posture réaliste à l'égard du modèle arborescent est entretenue par nombre de phylogénéticiens et de phylogénéticiennes. Cette posture se traduit notamment par la croyance selon laquelle l'évolution du monde biologique pourrait être représentée par l'Arbre du vivant, au singulier. Bien que le corpus disciplinaire soit plus ambivalent à l'égard des réseaux qui sont souvent considérés comme n'étant que des outils mathématiques utiles pour comprendre les relations entre organismes et gènes, les réseaux phylogénétiques sont tout de même parfois réifiés, comme je l'ai démontré plus tôt. Ainsi, une personne encline à adopter un réalisme à l'égard des théories, des modèles, des hypothèses ou des concepts scientifiques aurait donc, au sein de la phylogénétique, au moins deux cibles potentielles pour assouvir ses pulsions philosophiques : l'arbre et le réseau. Velasco (2012, 2013)

a exploré la question suivante : lequel des deux modèles devrait être considéré comme une représentation de la structure *réelle* des liens phylogénétiques entre unités taxonomiques?

L'analyse de Velasco est selon moi tout à fait adéquate et il est inutile de pousser la chose plus loin. En bref, son argument consiste à résumer l'ensemble des phénomènes qui ont motivé les critiques du modèle arborescent et le développement de modèles réseaux. Il souligne l'importance de l'endosymbiose, de l'hybridation des lignées et des transferts horizontaux de matériaux génétiques. Ces phénomènes bien documentés étant incompatibles avec le modèle arborescent (puisque'ils génèrent des lignées avec plus d'un ancêtre), le ou la réaliste n'a d'autre choix que de se tourner vers un modèle qui peut en rendre compte (si on présume que le ou la réaliste n'accepte comme réel que ce qui correspond aux phénomènes). Velasco conclut son analyse de la question comme suit : « The empirical evidence is clear; there are a great many nontreelike processes that produce nontreelike genealogical patterns in nature. The realist is forced to become a realist about networks rather than trees. » (Velasco 2012, 628).

Il pourrait cependant être argumenté que les différents phénomènes qui vont à l'encontre du modèle arborescent sont négligeables, en ce sens qu'ils ne font qu'obscurcir un patron phylogénétique sous-jacent qui est bien réel. DeSalle et Riley adoptent cette posture, comme déjà mentionné : « We want to make it very clear that we are not minimizing, ignoring or deflecting HGT as a biological process. It exists and it is an important factor in how many microbes have diverged and evolved. We simply suggest that it is not pervasive enough to fell a bifurcating tree of life. » (DeSalle et Riley 2020, §27). Pour défendre ce conservatisme, DeSalle et Riley évaluent quantitativement la présence de gènes ayant été transmis latéralement au sein de différentes lignées. Ces deux biologistes citent des travaux qui suggèrent que pour l'ensemble du monde vivant, jusqu'à 10 % des gènes auraient des histoires ponctuées de transferts horizontaux entre lignées (Zamani-Dahaj et al. 2016) et que ce taux pourrait monter jusqu'à 39 % dans le monde bactérien (DeSalle 2009). Bien que ces évaluations puissent paraître significatives, DeSalle et Riley considèrent qu'elles témoignent au contraire du fait que l'impact des transferts génétiques horizontaux est souvent surévalué (il et elle ne tiennent cependant pas compte d'autres phénomènes menant à la réticulation comme l'hybridation ou l'endosymbiose). Pour soutenir

leur conclusion, DeSalle et Riley soulignent la possibilité d'extraire un patron arborescent bien marqué malgré toute la transversalité observée :

In our view, the signal of vertical inheritance remains loud and clear. In the majority of cases where multiple genes and/or genomes have been used for phylogenetic reconstruction, the resulting trees are resolved and largely congruent. Furthermore—with some interesting exceptions—these reconstructions match established patterns of classification. (DeSalle and Riley 2020, §10)

Ainsi, DeSalle et Riley conservent l'idée d'une structure arborescente qui rendrait compte réalistement des relations phylogénétiques entre lignées.

C'est donc dire que, dans une posture réaliste, le choix entre les deux modèles disponibles dépend largement du degré de simplification ou d'abstraction qui est considéré comme étant acceptable. Il va sans dire que les modèles réseaux représentent eux aussi des simplifications à l'égard de l'objet d'étude. Il s'agit d'abstractions (ou d'idéalisations) qui, par leur capacité même à simplifier un objet complexe, le rendent plus accessible. Or, en enlevant une contrainte sur le champ des possibles (celle de la bifurcation), les modèles réseaux sont par définition plus complexes et mieux taillés pour faire écho à la complexité des liens phylogénétiques et de la réticulation aujourd'hui bien documentée (Velasco 2012). Nonobstant ce constat, une posture comme celle de DeSalle et Riley est considérée comme étant légitime par des praticiens et praticiennes qui accordent une suffisamment grande valeur à la convergence des données sur une topologie arborescente et qui considèrent que faire abstraction d'une partie des données est justifié.

En ce qui concerne l'analyse épistémologique de ces deux postures, il est impossible de faire prévaloir l'une des options puisque toutes deux, par leur réalisme, sont incompatibles avec les conséquences du problème de la coordination. En effet, les liens phylogénétiques sont inaccessibles à l'observation directe. Les pratiques de la phylogénétique visent à les inférer sur la base d'analyses de similarité et de modèles phylogénétiques (modèles arborescents ou réseaux). Ainsi, notre accès aux relations phylogénétiques est dépendant des modèles concernés, dont l'acceptation précède chaque observation (chaque itération repose sur l'acceptation d'un bagage théorique au sein duquel on retrouve le ou les modèles concernés). Si le réalisme à l'égard des modèles phylogénétiques se définit comme la croyance selon laquelle ces modèles (ou l'un

d'entre eux) sont des objets qui existent et dont l'existence est indépendante des croyances, pratiques linguistiques ou schèmes conceptuels de qui que ce soit (Miller 2019), alors force est d'admettre qu'il est tout simplement impossible de démontrer la chose. Jusqu'à preuve du contraire, toute démonstration de la validité d'un modèle dépend de pratiques scientifiques et donc de pratiques humaines. Postuler une existence indépendante revient donc à extrapoler au-delà de ce qui est empiriquement défendable.

En bref, ma posture face au réalisme (et face à l'antiréalisme qui *nie* la réalité d'un modèle, d'une théorie, d'une hypothèse, etc.), telle qu'elle est informée par le problème de la coordination et ses conséquences épistémiques, traduit une posture agnostique épistémologiquement justifiée à l'égard de ce qui existe indépendamment des pratiques scientifiques humaines. Au contraire, le réalisme implique des croyances non justifiées et non justifiables puisqu'elles portent sur des choses qui nous sont inaccessibles. Ma posture est loin de rejeter toute possibilité de normativité ou de progrès scientifique. Simplement, il s'agit de bien saisir la complexité des facteurs qui influencent ce progrès et qui régissent cette normativité. Une correspondance naïve entre la réalité telle qu'elle existe indépendamment de nous et les aspects symboliques des pratiques scientifiques n'a cependant pas sa place au sein de ce consortium de facteurs.

7.3.3 L'utilité respective des réseaux et des arbres

Puisque le problème de la coordination est sous-jacent à toute observation des relations phylogénétiques entre espèces, il est tout aussi difficile d'établir lequel des deux types de modèles est le plus utile pour les inférences phylogénétiques que de déterminer lequel des deux est le plus conforme à la réalité indépendante de notre cognition. En effet, l'utilisation d'un modèle implique que les données traitées reflètent ce modèle, d'une façon ou d'une autre. Cette situation a été soulignée par les praticiens et praticiennes. DeSalle et Riley (2020) suggèrent même que ce point a été utilisé pour miner la légitimité du modèle arborescent. En effet, cet argument peut être observé dans le corpus concerné : « Standard phylogenetic methods simply assume that the data are generated from a tree process and then find the best tree to fit the data. This is no way to detect whether a tree is the right kind of structure in the first place » (Velasco

2012, 630; voir aussi Doolittle et Baptiste 2007). Or, si cet argument est valable pour le modèle arborescent, il l'est aussi pour les modèles réseaux.

Ainsi, l'utilisation d'un modèle forge les résultats d'une manière telle que ceux-ci lui correspondent. Ce problème est doublé du fait que les modèles réseaux, étant plus complexes par définition, ne peuvent que mieux correspondre aux données (dans une approche centrée sur le maximum de vraisemblance) : « adding reticulations to a tree or network will monotonically improve the value of any optimality criterion based on the character data, because a more complex network can never fit the data worse than a simpler one (Makarenkov et Legendre 2000; Jin et al. 2007). » (Morrison 2014b, 632). Ces deux constats soulignent que l'utilité respective des deux modèles est difficile à départager, surtout si l'on considère que les deux sont mutuellement exclusifs.

En effet, même si un modèle donné correspondait avec exactitude et complétude à l'histoire évolutive des espèces, une diversité de modèles pourrait tout de même être utile pour appréhender cette histoire évolutive. Chaque modèle pourrait, par exemple, mettre l'accent sur certaines dimensions du processus évolutif et ainsi permettre aux communautés scientifiques d'utiliser un modèle approprié à chaque problématique. Ainsi, même si on considère que l'histoire évolutive est irrémédiablement réticulée, on peut en étudier certains phénomènes (par exemple l'évolution d'un segment d'ADN donné) sur la base de modèles arborescents; l'inverse est aussi vrai.

Cette diversification des tactiques est notamment nécessaire lorsque l'on cherche à croiser des études portant sur différents niveaux d'organisation. L'histoire des espèces, si elle est comprise comme l'histoire de collections de gènes, est éminemment réticulée, même chez les eucaryotes pluricellulaires dont les dynamiques évolutives sont caractérisées par des hybridations fréquentes (Mallet 2005; Mallet, Besansky et Hahn 2016; Ottenburghs 2019; Vanderpool et al. 2020). Cependant, de nombreux chercheurs et de nombreuses chercheuses considèrent que pour les organismes, cette histoire peut tout de même être efficacement traduite par un arbre. De surcroît, la diversité des tactiques demeure utile puisque la puissance calculatoire requise pour mobiliser les modèles réseaux est plus grande que celle nécessaire pour

l'utilisation de modèles arborescents (Morrison 2014a; Wen et al. 2018). L'analyse de grandes quantités de données gagne donc, dans certains cas, à être réalisée par l'entremise de modèles arborescents.

Considérons un exemple de pratiques mobilisant les deux types de modèles. Vanderpool et al. (2020) ont récemment exploré la phylogénie des primates en mobilisant un ensemble de séquences beaucoup plus grand que les études précédentes sur le sujet ne l'avaient fait (environ 25 fois plus de séquences ont été mobilisées que dans les autres études citées par les auteurs et autrices, soit Perelman et al. 2011 et Springer et al. 2012). En utilisant le même modèle que les études précédentes, soit le modèle arborescent, Vanderpool et al. ont obtenu une topologie presque identique à celle déjà considérée comme étant consensuelle. Cependant, le groupe de recherche ne s'est pas arrêté là. Bien que la topologie arborescente des primates (obtenue par des approches de maximum de vraisemblance et bayésiennes) témoigne d'une grande robustesse, il va sans dire que les arbres phylogénétiques des gènes constitutifs du jeu de données ne concordent pas tous les uns avec les autres (ils ne concordent donc pas tous sur la topologie des primates). Vanderpool et al. ont alors identifié des événements d'introgression (intégration de matériel génétique au sein du génome d'une lignée à la suite de l'hybridation) qui expliqueraient les conflits entre arbres génétiques.

Autrement dit, Vanderpool et al. ont cherché à offrir une représentation réticulée des gènes de primates qui est, selon eux et elles, compatible avec le patron arborescent qui traduit l'histoire des espèces de primates, les deux représentations offrant des informations distinctes mais complémentaires. Afin de détecter les événements d'introgression, Vanderpool et al. ont tenté de reconstruire des réseaux phylogénétiques en utilisant notamment PhyloNet, mais les résultats ont été peu concluants : « When there are multiple episodes of gene flow within a clade, even complex computational machinery may be unable to infer the correct combination of events. » (Vanderpool et al. 2020, 14). Ils et elles se sont donc tourné·e·s vers une méthode mise au point par Huson et al. (2005) permettant de localiser des événements d'introgression sur la base d'arbres non compatibles. Ainsi, bien que les arbres soient, au sein de cette étude, les principaux outils inférentiels, il n'en demeure pas moins qu'ils sont utilisés notamment pour représenter une histoire réticulée (comme en témoigne la quatrième figure de l'article).

Ainsi, malgré le fait que l'histoire évolutive soit réticulée, les arbres demeurent des outils intéressants. L'évaluation de l'utilité des réseaux doit ainsi être comprise en fonction des objectifs de recherche (Brigandt 2020) et des moyens disponibles pour les atteindre. Le cas d'étude tout juste exploré démontre que la négociation entre modèles rivaux peut très bien mener à l'amélioration des itérations passées au sein d'un champ d'études donné comme la systématique. Le cohérentisme historique invite en ce sens à considérer l'utilité des modèles scientifiques sur la base d'une analyse pragmatiste, interne aux pratiques contraintes par les conséquences épistémiques du problème de la coordination et nuancée. Cette approche contraste avec certains arguments présents au sein du corpus de textes étudié qui mobilisent une philosophie des sciences qui gagne à être dépassée.

Par exemple, toujours dans le même article de 2020, DeSalle et Riley (2020) campent une discussion relative à l'utilité du modèle arborescent en termes de vérificationnisme et de philosophie de la réfutabilité. Leur argument, en bref, consiste à démontrer que la phylogénétique traditionnelle (mobilisant le patron arborescent) n'est *pas* vérificationniste; elle est falsificationniste, et puisque les données disponibles ne réfutent *pas* le modèle arborescent, celui-ci devrait continuer à guider la systématique contemporaine⁴⁰. Doublé de la discussion sur la preuve intégrale du chapitre précédent et des récurrentes références à Popper, cet argument témoigne du fait que la phylogénétique, en tant que champ discursif, mobilise régulièrement des philosophies des sciences dont les apories ont préalablement été démontrées.

En effet, la philosophie de la réfutabilité et le vérificationnisme considèrent que l'empirie est le principal vecteur de légitimation des pratiques scientifiques (la philosophie de la réfutabilité est plus nuancée à ce sujet, mais ce n'est pas le cas du traitement que certains phylogénéticiens et certaines phylogénéticiennes lui réservent, comme démontré au chapitre précédent). Le débat relatif à l'utilité des modèles arborescents et réseaux est un parfait exemple du besoin de

⁴⁰ La philosophie de la réfutabilité et le vérificationnisme sont des approches qui permettent d'évaluer la scientificité et la valeur d'hypothèses, de modèles ou de théories. Dans le texte de DeSalle et Riley, les deux philosophies sont considérées comme des caractéristiques inhérentes ou potentiellement inhérentes à certaines pratiques scientifiques. Ici, je fais abstraction de cette inexactitude pour simplifier mon propos sans perdre l'essence de mon commentaire, soit que l'utilité des modèles étudiés ne peut être établie que sur la base d'une pluralité de facteurs.

considérer une multitude de facteurs pour comprendre les dynamiques scientifiques (tout comme le débat entourant *Methanonatronarchaeia* présenté au chapitre précédent). Doolittle, éminent biologiste de l'évolution, prêche en ce sens (le modèle des super-espèces mentionné dans la citation suivante est un modèle qui met de l'avant la réticulation évolutive; il s'agit d'un modèle réseau) :

This super-species model is at one end of a spectrum defined by the frequency and generality (in terms of genes that can be transferred) of lateral gene transfer. The traditional tree model is still at the other, *and there is no single experimental observation that could ever decide between them.* (Doolittle 1999, M7, italiques ajoutés)

Ce sont plutôt des principes théoriques (le fait que la contrainte topologique liée au principe de divergence soit justifiée ou non, par exemple) et des considérations pratiques (le fait que la puissance calculatoire nécessaire pour la reconstruction de réseaux phylogénétique est aujourd'hui disponible, par exemple), toujours en dialogue avec les observations, qui déterminent la teneur des débats et dynamiques de justifications scientifiques⁴¹. En ce sens, il est urgent pour la phylogénétique de mettre à jour son bagage philosophique afin d'acquérir des outils adéquats pour l'analyse réflexive des enjeux au cœur de ses pratiques contemporaines. Le cohérentisme historique offre une boîte à outils idéale pour ce faire.

7.3.4 De la place des lignées (et conséquemment de la phylogénétique) en biologie de l'évolution

En 7.3.3, j'ai souligné que l'utilité des modèles réseaux et arborescents dépend notamment des objectifs de recherche. Or, les deux types de modèles conviennent pour remplir les objectifs généraux de la discipline, soit la mise en relation d'unités taxonomiques opérationnelles qui permet notamment d'offrir un contenu empirique à la théorie de l'évolution par voie de sélection naturelle (voir chapitre 1 de la présente thèse). Afin de remplir ces objectifs,

⁴¹ Il convient ici de réitérer un principe méthodologique mentionné en introduction : mon travail d'analyse étant inscrit dans la tradition philosophique occidentale, je m'intéresse surtout à la teneur des discours qui structurent les arguments scientifiques. Anthropologues, sociologues, historiens et historiennes identifieraient indéniablement plus de facteurs que ceux mentionnés ici. Ce constat témoigne à la fois des limites de la présente analyse et de la portée unificatrice du cohérentisme historique.

la phylogénétique, qu'elle soit centrée sur les réseaux ou les arbres, met de l'avant une logique bien précise : il s'agit de retracer l'histoire du monde vivant, celle-ci étant réduite aux dynamiques de formation de lignées. Dans la présente section, je démontre que la décomposition d'un des amalgames présentés au chapitre précédent (selon lequel formation de lignées et évolution par voie de sélection naturelle seraient nécessairement liées) a d'importantes conséquences sur le rôle de la phylogénétique en biologie de l'évolution. Si cette critique est, à ma connaissance, absente du corpus concerné, il convient de noter que différentes complications liées à la notion de lignée, tout particulièrement au niveau des organismes, ont au contraire été soulevées par la communauté scientifique (par exemple, dans Boucher et Baptiste 2009).

Les pratiques phylogénétiques sont érigées sur l'idée selon laquelle l'histoire évolutive peut être décrite sur la base lignées de gènes, d'organismes, d'espèces, etc. De surcroît, la phylogénie est considérée comme le point de départ de toute biologie de l'évolution (voire de toute biologie) : « If 'nothing in biology makes sense except in the light of evolution' (Dobzhansky 1973), then surely phylogeny sets conditions for all biological study » (Kluge 2002, 591-592). Cette logique est traversée par la validité de la réduction de l'histoire évolutive du monde vivant à l'histoire des lignées. Or, cette réduction repose sur l'amalgame présenté au chapitre précédent et selon lequel l'évolution, et tout particulièrement l'évolution par voie de sélection naturelle, est *nécessairement* réalisée au sein de populations d'entités qui forment des lignées (Godfrey-Smith 2009; Griesemer 2000; Lewontin 1970; Wiley et Lieberman 2011). Cette croyance centrale de la phylogénétique est même conservée par des théoriciens et des théoriciennes ayant remis en question la nécessité de la reproduction pour qu'il y ait évolution par voie de sélection naturelle (Bouchard 2008; Bourrat 2014; Doolittle et Inkpen 2018). Ce n'est que très récemment que l'idée selon laquelle l'évolution par voie de sélection naturelle peut être indépendante de la formation de lignées a reçu l'attention qu'elle mérite, principalement sur la base d'analyses conceptuelles (voir chapitre précédent; Charbonneau 2014; Papale 2021; voir également Roughgarden 2020 pour une approche empirique). Sans traiter directement du rôle qu'occupe la formation de lignées au sein de l'évolution par voie de sélection naturelle, de nombreuses publications récentes portant sur la stabilité de traits biologiques indépendante des taxons qui les réalisent (Doolittle et Booth 2017; Doolittle et Inkpen 2018; Lenton 1998; Suárez 2020; Suárez et Triviño

2020; Taxis et al. 2015) et sur la nature phylogénétiquement composite des unités de sélection (Baptiste et Papale 2021; Ereshefsky et Pedroso 2015; Gilbert et Tauber 2016; Lloyd 2018; Papale, Saget et Baptiste 2020; Roughgarden 2020) renforcent la croyance selon laquelle l'explication des phénomènes évolutifs pourrait se faire sans explorer les dynamiques de formation de lignées. Pour le dire simplement, les relations phylogénétiques ne représentent qu'une fraction des interactions entre entités biologiques; les phénomènes évolutifs résultent de dynamiques entre entités qui sont des composites phylogénétiques plutôt que des tous unis pouvant former des lignées. De surcroît, comme détaillé au chapitre précédent, l'hérédité, condition nécessaire de l'évolution par voie de sélection naturelle, peut être réalisée au sein de populations dans lesquelles il n'y a pas de dynamiques de formation de lignées pertinentes ou au sein desquelles les lignées sont trop diffuses pour être causalement associées à la réalisation de l'hérédité (Papale 2021).

Prenons le cas de la résistance aux substances antimicrobiennes, de l'espèce bactérienne *Acinetobacter baumannii* et du gène *bla_{NDM}*. Malgré son apparition récente, ce gène est, à l'échelle mondiale, le plus commun parmi ceux contribuant à la résistance aux carbapénèmes, soit des antibiotiques à large spectre très utilisés par l'être humain (Logan et Weinstein 2017). Ce gène, qu'on retrouve au sein d'éléments génétiques mobiles (c'est-à-dire du matériel génétique facilement déplacé au sein d'un génome ou entre génomes), est probablement apparu au sein d'*Acinetobacter baumannii* bien que les occurrences de cette espèce bactérienne et du gène soient aujourd'hui largement indépendantes (Acman et al. 2021). De surcroît, *bla_{NDM}* lui-même est loin d'être un tout phylogénétique; il s'agit au contraire d'une chimère issue de la recombinaison de matériel génétique préexistant, une chimère s'étant construite au sein de *baumannii*, mais plus précisément au sein d'éléments génétiques mobiles où son existence est corrélée à celle d'autres séquences génétiques (Toleman et al. 2012). À toute cette complexité, il faut ajouter que la généralisation du gène *bla_{NDM}* au sein de la population est intimement liée à des conditions environnementales bien particulières, soit les pressions anthropogéniques liées à la production et à la consommation de carbapénèmes. Il a en effet été démontré que de telles pressions environnementales peuvent contribuer à la fixation des gènes au sein d'une population (Graham et al. 2016; Knapp et al. 2010).

Dans une telle situation, l'analyse phylogénétique ne peut offrir qu'un regard partiel sur le phénomène évolutif en question, soit la généralisation de la résistance aux carbapénèmes telle qu'elle est engendrée notamment par le gène *bla_{NDM}* (cette résistance peut être réalisée par d'autres mécanismes). Premièrement, le trait biologique ainsi généré (la résistance aux carbapénèmes) est un trait appartenant à des communautés microbiennes, des entités qui ne forment pas de lignées suffisamment univoques pour être objets d'étude de la phylogénétique. Pour que le gène soit exprimé de sorte à engendrer une résistance aux carbapénèmes, il faut qu'il soit intégré au sein de réseaux fonctionnels impliquant une variété de séquences génétiques (codantes ou non) et de structures cellulaires. Dans l'univers microbien, de tels réseaux fonctionnels dépassent les frontières des organismes et des taxons (Baptiste et Papale 2021; Ferreira et al. 2018). Ainsi, réduire l'histoire du phénomène de sélection qui implique un trait communautaire comme la résistance aux carbapénèmes à l'histoire d'une lignée génétique ou bactérienne occulte de fait le phénomène à l'étude en le simplifiant de façon non légitime. Il peut être utile de retracer les histoires généalogiques des composantes d'un système pour en comprendre l'origine, mais une discipline qui réduit l'histoire évolutive à l'histoire de ces lignées est *de facto* lacunaire : elle passe outre le phénomène évolutif d'intérêt.

Deuxièmement, cet exemple démontre bien la validité d'une affirmation faite ailleurs (Baptiste et Papale 2021; Papale, Saget et Baptiste 2020), soit que les entités traditionnellement étudiées par la biologie de l'évolution et la phylogénétique sont des composites. Dans le cas qui nous intéresse, l'unité fonctionnelle impliquée est la communauté bactérienne qui est un assemblage de pièces ayant des histoires évolutives distinctes (et plurielles). Réduire l'analyse à celle des taxons, en plus d'être trompeur, ne règle aucunement la situation. En effet, l'histoire du matériel génétique impliqué est indépendante de celle des taxons bactériens (*Acinetobacter baumannii* et *bla_{NDM}*, par exemple, ont des histoires évolutives indépendantes). Cette incapacité de la logique réductionniste de rendre compte de la situation est valide à différents niveaux : les taxons, les éléments génétiques mobiles et les gènes peuvent être et *sont*, dans bien des cas dont celui qui nous intéresse ici, des composites phylogénétiques, des chimères.

Le troisième élément du phénomène qui témoigne des limites explicatives de la phylogénétique est l'importance de l'environnement de sélection. Celui-ci, autant dans le

phénomène considéré que dans la théorie darwinienne, est le cœur du processus causal. L'environnement (qui inclut ici les pressions anthropogéniques) est au cœur des phénomènes d'évolution de deux façons : d'abord en tant que contexte de sélection; ensuite en tant que phénomène transformé par les populations qui évoluent. Le système en évolution est donc loin de former une population homogène d'entités dont on pourrait suivre les généalogies ascendantes ou descendantes. C'est un système de causalité multidirectionnelle et multiniveaux où les liens généalogiques sont bien présents, mais d'une importance devant être nuancée face à l'ensemble des interactions entre séquences génétiques, gènes, ensembles de gènes, protéines, structures cellulaires, cellules, organismes unicellulaires, organismes pluricellulaires, communautés d'organismes, écosystèmes locaux, écosystèmes globaux ou système planétaire. En d'autres termes, la phylogénétique enferme dans une boîte noire la majorité des phénomènes évolutifs, et notamment la sélection et la construction de niche, au profit de la formation de lignées. La validité de ce choix gagnerait à être réévaluée et des pratiques de rechange, à être développées. Si une réduction devait être opérée, ce ne serait pas celle de l'histoire évolutive à celle des lignées, mais bien la réduction de l'histoire des lignées à celle de leur contexte de sélection, principal vecteur explicatif des événements évolutifs.

Je tiens ici à bien préciser la portée de mon propos. Il est certainement utile de tracer des systèmes d'affinités entre entités biologiques à différents niveaux d'organisation et sur la base d'analyses de similarité. Il est aussi tout à fait possible que l'interprétation phylogénétique de ces systèmes et les ajustements qu'elle nécessite demeurent utiles même dans une perspective, comme celle que je propose, qui déconstruit le lien nécessaire entre étude de l'évolution et étude des lignées. Ce par rapport à quoi je suis critique est le rapport de la phylogénétique à ses objectifs généraux. D'une part, l'objectif qu'elle remplit efficacement, soit de retracer les lignées d'entités biologiques, me semble d'une importance qui mérite d'être nuancée dans le contexte de la biologie de l'évolution. D'autre part et en conséquence, la capacité de la phylogénétique à offrir un contenu empirique aux théories de l'évolution dépend d'un amalgame entre évolution et formation de lignées; la validité de ce complexe de croyances gagne à être réévaluée.

En résumé, la phylogénétique, avec ou sans réseaux, ne traduit que l'histoire simplifiée des entités pouvant former des lignées en faisant abstraction des mécanismes évolutifs qui

devraient être les cibles explicatives premières de la biologie de l'évolution : « Evolution is the hallmark of life. Descriptions of the evolution of microorganisms have provided a wealth of information, but knowledge regarding “what happened” has precluded a deeper understanding of “how” evolution has proceeded, as in the case of antimicrobial resistance. » (Baquero et al. 2021, e00050-19). Il y a donc un débat important à avoir : quelle devrait être la place de la phylogénétique au sein de la biologie de l'évolution?

7.3.5 La pensée populationnelle, la phylogénétique et le traitement de l'information biologique

L'objectif qui consiste à retracer l'histoire de lignées biologiques est loin d'être le seul point commun aux deux itérations phylogénétiques étudiées dans cette thèse. J'ai noté, en 7.2.3, un autre point de contact, soit que la phylogénétique, dans son ensemble, est une discipline classificatoire, c'est-à-dire une discipline qui décrit le monde par découpage et discrimination et en mettant en relation les catégories ainsi délimitées. Comme la prévalence de la nomenclature linnéenne en témoigne, les descriptions classificatoires ancrées dans l'analyse phylogénétique débordent généreusement de cette discipline : les descriptions des organismes sur la base de leur catégorie d'appartenance dominant la biologie. Dans la présente section, j'argumente que la biologie de l'évolution gagnerait à prioriser des méthodes descriptives quantitatives⁴². Le point central de mon argument est que les descriptions classificatoires excluent de leur propos les particularités des individus qui sont pourtant sources d'informations cruciales, surtout en contexte d'évolution.

Les groupes taxonomiques étudiés en phylogénétique sont représentés par des données issues d'un ou de peu de génomes. Par exemple, Vanderpool et al. (2020) reconstruisent la phylogénie des primates sur la base d'alignements de séquences où chaque séquence (qui est

⁴² La phénétique, mouvement taxonomique de la deuxième moitié du 20^e siècle, a largement contribué au développement de méthodes d'analyse de la similarité dans une perspective quantitative. Le rejet maintenant généralisé de ce courant de pensée (et non des méthodes qui y ont été développées) est le produit de l'ambition des phénéticiens et des phénéticiennes de développer une taxonomie qui serait dénudée de toutes contraintes théoriques. Un tel objectif est ici hors de question. Il s'agit d'octroyer aux descriptions quantitatives de la biodiversité une place centrale, mais le tout relève de pratiques théoriquement chargées; que cette charge théorique soit phylogénétique ou autre, son importance est indiscutable.

une concaténation de séquences génétiques codant pour des protéines) représente une espèce dans son ensemble. Ainsi, le portrait génétique d'un spécimen (lui-même réduit à la séquence concaténée) est utilisé pour établir les relations phylogénétiques de cette espèce avec le reste des primates. Au-delà de cette situation précise, le processus épistémique est malléable et dépend du niveau d'analyse des études phylogénétiques. Plusieurs spécimens d'une même espèce peuvent être considérés, notamment lorsque le but est de retracer les liens phylogénétiques entre populations au sein de cette espèce (Skerratt et al. 2002, par exemple). Lorsque des clades plus inclusifs comme *Methanonatronarchaeia* sont étudiés (Sorokin et al. 2018; Aouad et al. 2018), chacun peut être représenté par un plus grand nombre de séquences (dans le cas d'analyses de microbes, le génome est souvent séquencé à partir d'une population de clones), ce qui permet une meilleure analyse des affinités entre le clade cible et d'autres clades inclusifs, comme *Halobacteria*. Un point méthodologique demeure fixe malgré cette flexibilité : les séquences utilisées au sein des alignements représentent des populations entières.

Cette logique permet d'étudier les liens phylogénétiques entre groupes taxonomiques en tenant pour acquises les nombreuses similarités qui lient indéniablement les membres inclus dans une unité taxonomique. Nonobstant la validité de ce constat, il est autant indéniable que la différence entre les membres d'une espèce est évacuée du processus au profit d'une représentation simplifiée de populations entières. Il est tout aussi indéniable qu'il s'agit d'une source d'information extrêmement utile dans le contexte de l'étude de l'évolution biologique. Après tout, la différence entre les membres d'une population est le cœur du phénomène évolutif mis à jour par Darwin.

En effet, la théorie de l'évolution par voie de sélection naturelle est centrée sur les différences entre individus. Ces différences, lorsqu'elles engendrent des différences de fitness, permettent à certaines variantes d'accroître leur présence au sein d'une population, et au fil des générations, s'accumulent pour mener à des adaptations. C'est donc dire que Darwin a exhorté les théoriciens et théoriciennes de son époque à accorder une place à ces différences individuelles au sein d'une biologie qui avait jusqu'alors été largement dominée par l'étude des catégories et des archétypes les représentant. Cette contribution darwinienne, popularisée par l'entremise de l'expression *pensée populationnelle* (présentée au chapitre précédent), est cruciale pour le

développement d'une pensée de l'évolution comme celle que l'on connaît aujourd'hui : en traitant les différences individuelles comme étant accidentelles (ce qui était la norme à l'époque prédarwinienne), on néglige de fait un élément central de l'évolution par voie de sélection naturelle, soit la *différence* de fitness entre individus. Un mode descriptif qui escamote les particularités des différentes occurrences est donc mal adapté à l'étude de l'évolution si celle-ci est à comprendre dans un cadre darwinien (sans être réduite pour autant à l'influence de la sélection naturelle sur les dynamiques évolutives; voir Watson, Lopez et Baptiste 2021 pour une critique similaire du traitement phylogénétique des différences entre les individus).

Loin de moi la suggestion que la biologie, dans son ensemble, passe outre les différences interindividuelles. Cependant, la phylogénétique et le rôle de base qu'on lui assigne (voir notamment la citation de Kluge 2002 donnée au début de la section 7.3.4) témoignent de la prépondérance de la classification parmi les modalités descriptives. À ce point-ci, il pourrait être argumenté que la phylogénétique est une discipline plutôt historique que classificatoire (Cleland 2002; Lecointre et Huneman 2020; O'Malley 2016; Velasco 2012), ce qui ajouterait un sérieux bémol aux commentaires critiques qui précèdent. Une dimension historique est effectivement inaliénable dans le contexte phylogénétique, les systèmes de relations entre unités taxonomiques étant ultimement interprétés comme des séquences de divergence (et de réticulation) des lignées, comme une succession d'événements ayant eu lieu à des moments précis au sein de l'histoire du monde vivant.

Cette dimension historique se superpose toutefois à la dimension classificatoire du projet. En effet, l'interprétation historique de la mise en relation d'unités taxonomiques opérationnelles demeure secondaire au travail effectué qui, comme je l'ai souligné aux chapitres 1, 5, 6 et plus tôt dans celui-ci, est principalement composé d'analyses de similarité. Les résultats de ces analyses sont utilisés pour former des systèmes (arborescents ou réticulés) dits phylogénétiques. La dimension historique de ces systèmes *peut* être confirmée ou infirmée par la mobilisation de données extérieures issues, par exemple, de datation de fossiles. Parmi les différents témoignages du fait que la dimension historique est secondaire à la dimension classificatoire, on retrouve la mobilisation fréquente de réseaux ou d'arbres non enracinés pour représenter les affinités entre taxons. Ces représentations non enracinées mettent de côté la dimension

historique puisqu'il s'agit de réseaux (au sens de la théorie des graphes) non dirigés et abstraits. Elles demeurent d'excellents outils pour explorer les liens entre les unités taxonomiques, mais la question se pose : s'agit-il bien de représentations phylogénétiques, si la phylogénétique est une science historique? Cette question reste ouverte, mais son équivalent classificatoire l'est moins : la construction de tels systèmes est indubitablement classificatoire puisque des catégories sont mises en relation.

C'est donc dire que les pratiques classificatoires dominent la phylogénétique, et notamment par l'entremise de la nomenclature linnéenne, s'imposent sur une importante partie des pratiques biologiques. Cela implique que les relations entre individus sont considérées comme étant accidentelles et négligeables alors même que le cadre darwinien nous exhorte à mettre ces différences au cœur de notre champ d'études. Suivre la logique de la pensée populationnelle jusqu'au bout revient à exclure les descriptions classificatoires de la biologie de l'évolution. Sans vouloir soutenir une posture aussi forte (la classification a minimalement une fonction communicationnelle en simplifiant la référence aux phénomènes), j'entends tout de même suggérer qu'un tournant quantitatif permettrait un renouveau du regard posé sur la biodiversité.

Une manière d'envisager la chose serait de faire de la notion d'homologie un concept graduel (comme la température ou le temps). Les homologies, particulièrement au niveau moléculaire, sont identifiées sur la base de modèles (comme présenté au chapitre 5) qui permettent de pondérer les différences entre séquences étudiées. Or, dans la pratique, cette quantification de la similarité s'efface face au traitement de l'homologie dans une logique « tout ou rien » : soit deux séquences sont homologues, soit elles ne le sont pas. Cette discrimination est effectuée sur la base de seuils de similarités (variables) qui minimisent l'importance de la quantification (dans la citation suivante, l'utilisation du terme *qualitative* [qualitatif] porte à confusion; il gagnerait à être remplacé par *discrete* [discrète] ou *categorical* [catégorique]) : « Thus, the consensus among molecular biologists became that similarity was defined as quantitative by comparing the sequences in questions, but that homology was qualitative – sequences are homologs or they are not » (Haggerty et al. 2014, 503).

En contraste, un traitement graduel de l'homologie décrirait mieux les affinités entre traits biologiques (Haggerty et al. 2014; Rieppel 2015; voir aussi West-Eberhard 2003, mais l'homologie y est considérée dans une perspective développementale). Plutôt que de déterminer les homologies sur la base d'une similarité qui dépasse ou non un certain seuil, il serait question de déterminer à quel point deux séquences (ou autres traits) sont homologues. La notion d'homologie étant le cœur inférentiel des pratiques phylogénétiques, celles-ci seraient remodelées en conséquence. Spécifiquement, les homologies témoigneraient de liens phylogénétiques plus ou moins forts entre une pluralité d'entités et non de séquences univoques de divergences successives (l'utilisation des réseaux de similarité de séquences travaille en ce sens, mais les séquences génétiques y représentent toujours des populations entières; voir notamment Watson et al. 2019).

L'impact d'une telle transition serait plus nuancé qu'il peut le sembler au premier regard. Prenons une tentative récente de standardiser l'usage des différents rangs taxonomiques à l'échelle bactérienne (Parks et al. 2018). Cette entreprise colossale vise à retracer les affinités, dans ce cas-ci les liens phylogénétiques, au sein du monde microbien, puis à utiliser les résultats pour déterminer quels groupes monophylétiques doivent être considérés comme des espèces, des genres, des familles, des ordres, des classes, etc. (une entreprise classificatoire s'il en est une). Parks et al. en viennent à proposer des ajustements significatifs à la taxonomie bactérienne (58 % des 94 759 génomes étudiés ont vu leur étiquetage taxonomique modifié). Ce qui est intéressant, pour mon propos, est que cette réorganisation de la taxonomie bactérienne repose sur une quantification des affinités entre lignées, soit le calcul de la divergence évolutive relative (*relative evolutionary divergence* ou *RED*). Parks et al. assignent les différents rangs taxonomiques aux clades sur la base de cette mesure. Le calcul de la divergence relative des entités biologiques, décrit dans Parks et al. 2018, requiert des points de référence fixes (l'ancêtre commun à l'ensemble des bactéries dont la valeur de la divergence relative est fixée à 0 et celle des phylums terminaux fixée à 1) et la reconstruction d'une topologie arborescente (le calcul de la divergence relative implique notamment une référence à l'ancêtre proximal d'un phylum donné). Ce que je propose est l'utilisation d'une telle mesure pour décrire les organismes directement, sans qu'elle soit par la suite associée à des rangs taxonomiques. Dans une perspective phylogénétique, une

mesure comme la divergence relative pourrait être utilisée pour décrire non seulement des unités taxonomiques, mais aussi des individus, des spécimens et leurs particularités.

Dans un contexte *non* phylogénétique, une mesure similaire pourrait être mise sur pied autrement qu'à partir de topologies phylogénétiques. Les réseaux de similarité de séquences (Watson et al. 2019), qui présentent la similarité en insistant sur les affinités pluridirectionnelles entre séquences génétiques ou protéiques, sont d'excellents points de départ pour une telle entreprise (points qui pourraient être modulés pour l'analyse d'autres niveaux d'organisation biologique). Les arêtes entre nœuds peuvent être pondérées, au sein de ce type de réseaux, sur la base des scores des similarités qui lient deux séquences (faisant ainsi écho à la conception graduelle de l'homologie dans un contexte non phylogénétique). La similarité entre deux nœuds liés indirectement pourrait aussi être quantifiée sur la base des chemins, au sein du réseau, qui lient ces deux nœuds (le nombre de nœuds les séparant pourrait être utilisé comme métrique, par exemple). D'autres paramètres topologiques pourraient aussi être mobilisés et quantifiés, comme le degré des nœuds, la présence (quantifiable) de motifs constitutifs des réseaux, leur diamètre, la connectance de l'ensemble du système étudié, etc.

C'est donc dire que la description de la biodiversité du monde biologique pourrait être basée sur différents paramètres, que ceux-ci concernent les entités biologiques auxquelles nous sommes habitué·e·s (gènes, organismes, taxons, familles de gènes, etc.) ou des entités plus complexes représentées sous forme de réseaux (Baptiste et Papale 2021). La théorie des graphes pourrait être utilisée pour explorer ces réseaux et ainsi tenter de comprendre ce qui fait des réseaux biologiques des réseaux non aléatoires (Watson, Habib et Baptiste 2020) tout en priorisant la quantification des entités biologiques plutôt que leur classification.

L'établissement de mesures standardisées qui permettraient de décrire les entités biologiques de manière constructive, et éventuellement d'établir des modalités descriptives consensuelles comparables à celles inscrites dans la nomenclature binomiale, est très certainement un défi immense pour les biologistes. Malgré les importantes difficultés qui jonchent la voie vers la priorisation des descriptions quantitatives en biologie, il y a au moins deux arguments qui incitent à croire que le jeu pourrait en valoir la chandelle. Premièrement, comme

exploré plus tôt, la transition aux descriptions quantitatives permet d'intégrer au sein des descriptions de la biodiversité les différences entre individus, une source d'information importante occultée par les descriptions classificatoires.

Deuxièmement, l'histoire de la tradition scientifique occidentale nous apprend qu'une telle transition peut effectivement être porteuse de progrès épistémiques. La mathématisation du mouvement des corps, en contraste avec l'explication ptoléméenne de ce mouvement ancrée dans la *nature* des corps (le mouvement de tous les corps de même nature se comprend par leur appartenance à la catégorie concernée; par exemple, tous les corps célestes ont une vitesse constante et un mouvement circulaire), s'est révélée cruciale pour le développement de la physique :

Face aux révélations de la lunette, par exemple les montagnes de la lune ou les taches du soleil, la réaction normale d'un tenant de cette philosophie [précopernicienne] était d'en chercher une explication (en l'occurrence une neutralisation) en invoquant l'essence ou la nature propre de ces corps. Or c'est là supposer non seulement qu'il y a des essences [ou une nature propre aux corps], mais que nous en avons une connaissance directe, et que grâce à elle nous pouvons rendre compte des données de l'observation. Pour l'astronome-philosophe, habitué au contraire à raisonner à partir de ces données, une telle démarche est à la fois illusoire et stérile. (Clavelin 2004, 366)

Ainsi, les phénomènes physiques ont longtemps été expliqués par l'entremise de spéculations sur la nature des corps, une pratique renversée par la révolution copernicienne. Par quoi ces réflexions d'ordre classificatoire ont-elles été remplacées?

Le renversement est complet au regard de la situation traditionnelle où seule la cosmologie philosophique [spéculant sur la nature des corps] avait compétence pour traiter des questions physiques et où la géométrie se bornait à 'sauver les phénomènes'. De subordonnée, le rôle de la géométrie est clairement devenu essentiel. (Clavelin 2004, 368)

Que nous ayons un accès épistémique aux dynamiques de formation de lignées ou non, la question se pose, par analogie : décrire les entités sur la base de leur clade d'appartenance est-elle la méthode descriptive la plus appropriée dans un contexte d'évolution? Des analyses quantitatives de la similarité me semblent plus aptes à rendre compte des informations importantes. Ce point de vue, j'en suis bien conscient, s'inscrit en opposition avec la

compréhension habituelle des pratiques classificatoires et de leur rôle en biologie. Considérons le passage suivant d'un texte récent portant sur la classification biologique :

Si notre langage était réglé au niveau de description le plus près possible de cette historicité [qui est responsable des particularités de chaque entité], il faudrait faire de la biologie en donnant un prénom à chaque mouche, à chaque pâquerette. Pire, chaque cellule étant différente de sa voisine, nous en viendrions à donner un prénom à chaque cellule. Nous avons pourtant besoin de concepts et de noms de portée générale, ce qui rend le problème de la ressemblance particulièrement saillant en biologie : la généralité, et les mots qui s'y rattachent, se construisent sur le constat de la ressemblance. [...] Le présent article remet en perspective une contrainte rationnelle profonde, bien que trop souvent oubliée : d'une manière générale en science, nos concepts ne sont rationnellement efficaces que si nous classons les entités du monde sur la base de partages de propriétés [...]. Ensuite, plus spécifiquement en sciences biologiques, ces partages traduisent un focus sur l'ascendance commune posée comme réquisit par Darwin en 1859, et rendu réalisable par Willi Hennig en 1950. (Lecointre et Huneman 2020, 77)

Cette citation résume l'ensemble des croyances que la biologie, centrée sur les pratiques classificatoires, soutient de façon implicite et quelques fois, comme dans cet article, de façon explicite. S'il est extrêmement utile de classer pour simplifier la référence aux phénomènes (et éviter d'avoir à baptiser chaque cellule), il faut se rappeler que cette classification, d'un point de vue informatif, est souvent triviale. Le terme « planète », par exemple, est utile d'un point de vue référentiel, mais il est inutile lorsqu'il est temps de comprendre le mouvement des corps. Sinon, en effet, comment pourrions-nous comprendre les différences entre le mouvement de deux planètes distinctes? En biologie de l'évolution, les concepts classificatoires comme ceux d'espèce ou de famille de gènes ne jouent cependant pas qu'un rôle référentiel; ils sont au cœur de notre compréhension de l'évolution. La structure inférentielle de la phylogénétique en témoigne bien. Il s'agit d'utiliser une séquence pour représenter des populations entières, pour ensuite mettre ces populations entières, considérées comme des tous phylogénétiques, en relation entre elles selon un seul critère, soit le critère généalogique. Cette logique tait de nombreuses différences qui promettent pourtant d'être source d'importantes informations relatives à l'évolution. Ainsi, la « contrainte rationnelle » de Lecointre et Huneman qui impose la classification est peut-être « profonde », mais elle représente aussi une perte d'information notable.

De surcroît, la dernière phrase de la citation exemplifie l'amalgame selon lequel sélection naturelle, lignées et, les auteurs ajoutent, classification seraient des notions liées de manière

nécessaire. Or, j'ai démontré de façon efficace que sélection naturelle et formation de lignées sont des idées indépendantes, et que, de toute façon, la divergence des entités biologiques pourrait être décrite sur le mode de la quantification plutôt que sur le mode classificatoire. Bien que l'impératif classificatoire soit enraciné profondément au sein de la culture intellectuelle de la biologie occidentale (comme le démontre l'article tout juste cité), la perte d'information inhérente à ce type de descriptions et la disponibilité de méthodes de rechange justifient d'envisager sérieusement un renouveau de notre rapport descriptif à la biodiversité.

7.4 Conclusion

L'objectif de ce chapitre était de réaliser une analyse comparative de la phylogénétique traditionnelle et de la phylogénétique centrée sur les modèles réseaux. Après avoir présenté les pratiques centrées sur les réseaux, j'ai réalisé une telle comparaison en étudiant certains enjeux auxquels toute itération phylogénétique est nécessairement confrontée. J'ai noté des différences majeures entre les modèles réseaux et arborescents en ce qui concerne la représentation des données et tout particulièrement de celles témoignant de réticulation au sein des dynamiques évolutives. Ces données sont traitées comme du bruit qui doit être surmonté par la phylogénétique traditionnelle, alors que celle centrée sur les réseaux est motivée par le besoin de tenir compte de ces données pour offrir un regard plus complet sur l'histoire du monde vivant. Cela témoigne d'une remise en question de la priorité de la divergence au sein du cadre théorique mobilisé.

Certaines différences plus nuancées ont aussi été notées. Les approches réseaux tendent à présenter des postures moins réalistes à l'égard des modèles que la posture traditionnelle ne l'est à l'égard de l'Arbre du vivant. Les approches réseaux, en plus de mobiliser des modèles réticulés, utilisent les outils arborescents de l'itération traditionnelle et considèrent que les meilleurs résultats (toutes choses étant égales par ailleurs) sont ceux qui présentent le moins de réticulation.

J'ai aussi noté des similarités marquantes entre les différentes itérations de la phylogénétique qui entretiennent une continuité, source de légitimité des itérations plus récentes. Dans une perspective abstraite, il convient en effet de souligner que la structure

inférentielle demeure la même dans les deux cas : des analyses de similarité et l'organisation des liens de similarité au sein d'un système (arborescent ou réticulé) pouvant être interprété phylogénétiquement représentent le cœur de l'entreprise. Cet important point de contact est renforcé par le fait que l'utilisation de modèles réseaux et arborescents repose sur les mêmes méthodes d'analyse de la similarité. Ces analyses de similarité sont les éléments les plus stables de la systématique. Même si les itérations phylogénétiques de la discipline en venaient à être surpassées (si les critiques mentionnées en 7.3.4 et 7.3.5 en venaient à être acceptées, par exemple), il faudrait s'attendre à ce que les analyses de similarités, et surtout les méthodes développées pour analyser la similarité, continuent d'être impliquées dans la spirale qui caractérise l'évolution de la systématique (au même titre que le rejet de la phénétique s'est effectué tout en intégrant les méthodes d'analyse phénétiques à la cladistique et plus généralement à la systématique phylogénétique).

Dans la deuxième partie du chapitre, j'ai appréhendé certains débats philosophiques en lien avec les pratiques phylogénétiques récentes. L'analyse de ces débats a été informée par l'analyse comparative réalisée préalablement. J'ai d'abord exploré l'affirmation, présente au sein du champ disciplinaire étudié, selon laquelle les réseaux phylogénétiques seraient des représentations incompatibles avec le cadre théorique darwinien. Au-delà du fait que cette affirmation est inexacte, j'ai profité de l'occasion pour souligner l'importance d'un modèle épistémologique comme celui du cohérentisme historique pour analyser de telles questions. En effet, le fait même de suggérer que l'utilisation d'un modèle, que la réticulation ou que l'accumulation de données massives issues de la génomique puissent remettre en question, en bloc, l'ensemble du cadre darwinien relève d'une conception de ce cadre qui repose sur des amalgames que le cohérentisme historique aide à décortiquer.

Les deuxième et troisième débats relèvent du besoin de trancher entre les deux modèles. Dans une perspective réaliste, les modèles réseaux doivent être considérés comme étant plus adéquats puisqu'ils rendent compte de phénomènes bien documentés, comme les transferts génétiques latéraux, que le modèle arborescent peine à gérer. Après avoir présenté ce constat en 7.3.2, je réitère ma posture par rapport à la question du réalisme (et de l'antiréalisme). Je défends l'idée selon laquelle le problème de la coordination implique que toute observation dépend d'un

cadre théorique et donc que nous n'avons aucun accès à la nature telle qu'elle existe indépendamment de ces cadres théoriques. De ce fait, les affirmations relatives à la conformité des connaissances scientifiques à une réalité indépendante de notre cognition (ou à l'absence d'une telle conformité) vont au-delà de ce qui peut être défendu empiriquement par rapport aux phénomènes, soit les entreprises de construction de connaissances scientifiques.

En 7.3.3, j'ai considéré l'utilité respective des modèles réseaux et arborescents. L'utilité des deux types de modèles est bien établie et les deux témoignent d'une certaine complémentarité. En ce sens, les deux gagnent à être utilisés lorsque nécessaire, ce qui évite d'avoir à trancher quant à la supériorité épistémique de l'un ou de l'autre, une tâche ardue considérant le problème de la coordination.

Finalement, j'ai soulevé deux débats qui gagnent à faire l'objet d'une attention toute particulière, bien qu'ils soient absents du corpus de textes étudié. Ces deux débats concernent la place centrale de la phylogénétique au sein de la biologie de l'évolution. Premièrement, je remets en question cette centralité en démontrant que l'évolution est un phénomène qui est irréductible à l'histoire de lignées verticales ou quasi verticales. Les phénomènes de formation de lignées, même quand ils sont présents, sont souvent trop diffus pour être causalement significatifs (par exemple, la récurrence de communautés microbiennes implique la formation de lignées, mais la teneur pluridirectionnelle de ces relations de parenté empêche l'identification de patrons univoques requis par les modèles phylogénétiques). L'étude de l'évolution gagne donc à se distancier de la notion de lignée, et incidemment, à être recentrée sur l'étude de dynamiques autres que généalogiques.

Deuxièmement, j'ai souligné un problème profond de la phylogénétique qui découle du fait qu'il s'agit d'une discipline classificatoire. La phylogénétique a comme objet d'étude des unités taxonomiques qui sont représentées et analysées par l'entremise d'une ou d'un petit nombre d'occurrences. Autrement dit, l'étude d'un spécimen type (ou de quelques-uns, mais toujours d'un nombre limité) suffirait à comprendre l'ensemble de la population qu'il représente. Les particularités des occurrences individuelles sont ainsi reléguées à l'arrière-plan du projet phylogénétique et de l'ensemble de la biologie qui mobilise la nomenclature linnéenne pour

appréhender les entités biologiques. Afin de pallier cette perte d'information inhérente aux projets classificatoires, j'ai souligné la capacité des méthodes quantitatives à gérer les différences interindividuelles activement. Des méthodes phylogénétiques pour mesurer la biodiversité sont déjà disponibles, mais rien ne nous force, ultimement, même dans un cadre darwinien, à décrire le monde vivant en priorisant les dynamiques de formation de lignées.

Conclusion de la thèse

L'objectif principal de cette thèse était de développer une approche philosophique pour l'analyse des pratiques classificatoires en sciences. En ce sens, j'ai démontré que le cohérentisme historique, un courant développé en philosophie de la mesure, est tout aussi efficace pour l'analyse des pratiques classificatoires que pour l'analyse des pratiques métrologiques. J'ai illustré les apports de cette approche philosophique par l'analyse des pratiques phylogénétiques contemporaines, soit des pratiques de classification centrales à la biologie de l'évolution. Plus spécifiquement, j'ai mené une analyse comparative de deux ensembles de pratiques actuellement présents au sein de cette discipline : les pratiques centrées sur les modèles arborescents et celles centrées sur les modèles réseaux. À la lumière du cohérentisme historique, cette analyse comparative m'a permis de décrire finement les similarités et différences entre ces deux approches phylogénétiques ainsi que de bien saisir les enjeux épistémiques de la discipline.

L'ensemble de ces démarches a été réalisé en deux étapes. La première étape de la thèse (chapitres 1 à 4) a exposé les principaux éléments de philosophie des sciences nécessaires à la bonne compréhension du courant de pensée que j'appelle le cohérentisme historique et du rôle que je lui assigne au sein de la philosophie de sciences. La deuxième étape de la thèse (chapitres 5 à 7) a mobilisé les outils issus du cohérentisme historique afin d'analyser les pratiques phylogénétiques contemporaines. Dans ce qui suit, je reviens sur le contenu de chacun des chapitres tour à tour.

Le premier chapitre de la thèse avait une double mission. D'abord, il s'agissait de présenter le problème de la coordination, un phénomène épistémique à la base de mon projet doctoral. Ensuite, j'ai démontré que ce problème s'incarne effectivement au sein de la structure inférentielle de la phylogénétique. La coordination est ce processus par lequel la dimension symbolique des pratiques scientifiques, ce qui inclut notamment des quantités théoriques ou des relations entre catégories d'objets, est mise en relation avec des aspects concrets des phénomènes. Cette mise en relation permet de fournir un contenu empirique aux théories, aux

modèles et aux hypothèses, ceux-ci demeurant de pures abstractions dépourvues de pertinence empirique tant qu'une telle coordination manque à l'appel.

La coordination est réalisée par la mise en place de procédures d'observation. Par exemple, en phylogénétique, des procédures permettent de lier des relations observables⁴³ au sein de la biodiversité (par exemple, des liens de similarité morphologique entre organismes) avec des relations inaccessibles à l'observation directe (comme l'ascendance commune). Ainsi, les liens phylogénétiques, malgré le fait qu'ils ne soient pas accessibles à l'observation directe (puisque nous ne pouvons pas observer le passé), peuvent être coordonnés avec des phénomènes biologiques, avec la biodiversité dans son ensemble.

La validité du lien inférentiel entre similarité et phylogénie ne pourrait toutefois être confirmée, dans l'absolu, que si des procédures étaient disponibles pour observer phylogénie et similarité *indépendamment* l'une de l'autre. Or, de telles procédures ne sont pas disponibles puisque l'objectif de la démarche est justement de mettre en place des procédures pour l'observation des liens phylogénétiques. Il faut donc accepter *de facto* un lien entre similarité et généalogie à l'échelle de l'ensemble du monde vivant. Ce lien étant théoriquement bien ancré au sein du cadre darwinien traditionnellement véhiculé, la situation témoigne concrètement de la charge théorique nécessaire à la coordination des théories et des phénomènes, une coordination rendue possible par l'observation empirique. *Ipsa facto*, la phylogénétique semble avoir, du moins au premier regard, une structure justificative circulaire (et en ce sens, il y a bien *problème* de la coordination) : l'observation des liens phylogénétiques, nécessaire à la corroboration de la théorie de l'évolution par voie de sélection naturelle, requiert d'accepter le lien phylogénie-similarité, un lien central à la théorie darwinienne de l'évolution. Or, comme je l'ai démontré en détail au premier chapitre de la thèse, par une discussion portant sur la notion d'homologie, cette circularité est artéfactuelle, c'est-à-dire qu'elle n'apparaît que lorsque des analyses épistémologiques sont réalisées dans une perspective anhistorique.

⁴³ En ce sens que des procédures d'observation sont déjà acceptées pour appréhender de telles relations.

L'homologie est la cheville ouvrière des inférences phylogénétiques puisqu'elle incarne le lien similarité-phylogénie. En effet, dans un cadre de pensée évolutionniste, l'homologie est la relation entre traits d'organismes distincts (ou d'unités taxonomiques distinctes) dont les similitudes sont le fruit d'une ascendance commune. En observant la mobilisation du concept à partir du début du 19^e siècle jusqu'à la mise en place de la cladistique, dans la deuxième moitié du 20^e siècle, il devient clair que les pratiques de la systématique, qu'elles soient phylogénétiques ou non, tirent leur légitimité des itérations passées de la discipline. En ce sens, le problème épistémologique de la coordination, qui est celui qui consiste à rendre observable ce qui ne l'est pas directement, mène à une structure de justification (en sciences) qui est hélicoïdale plutôt que circulaire. C'est bien la prise en compte explicite de la dimension diachronique de la structure épistémologique des pratiques scientifiques qui transforme le mouvement circulaire en hélice. Considérant que cette situation est au cœur des pratiques phylogénétiques, il convient, dans le but d'étudier épistémologiquement ces pratiques, de se munir d'un cadre philosophique qui permet de prendre en compte cette dimension diachronique des structures de justification phylogénétiques. Le deuxième chapitre de la thèse répond à ce besoin en présentant le courant de pensée que j'appelle le cohérentisme historique.

Ce courant de pensée regroupe un corpus de textes issus de la philosophie de la mesure (Chang 2004; Mari 2005; Morrison 1998; Shapin et Schaffer 1985; Tal 2016; van Fraassen 2008) et qui permettent d'appréhender les conséquences épistémologiques du problème de la coordination. L'objectif de ce chapitre était de démontrer que ce cadre philosophique permet d'intégrer les conséquences du problème de la coordination au sein d'une analyse épistémologique des pratiques classificatoires en sciences. J'y ai décrit ce courant de pensée en détail, après avoir exploré les courants philosophiques du 20^e siècle qui ont appréhendé le problème de la coordination (en limitant toutefois leurs propos à son incarnation métrologique). J'ai principalement présenté les approches fondationnalistes à l'épistémologie qui, en philosophie de la mesure, se sont manifestées par l'entremise du conventionnaliste (Mach 1896; Poincaré 1898; Reichenbach 1927) et les approches cohérentistes, représentées par le constructivisme (Tal 2012; Latour 1987; 2003). J'ai ainsi identifié les points forts et les apories de chacune de ces approches. En bref, le conventionnalisme met de l'avant l'importance de pouvoir rendre compte

du progrès scientifique, mais au prix d'une épistémologie anhistorique qui minimise les conséquences épistémiques du problème de la coordination. Le constructivisme, au contraire, cherche à tenir compte des difficultés liées à la coordination en mettant l'accent sur les déterminants non empiriques des pratiques scientifiques. Selon les critiques du constructivisme, cette importante contribution se fait toutefois au profit de la possibilité de rendre compte efficacement de la normativité et du progrès scientifique.

Le cohérentisme historique, que j'ai principalement exploré à partir des travaux de Chang (2004), de Tal (2016) et de van Fraassen (2008), intègre les forces du constructivisme et du conventionnalisme sans reproduire leurs faiblesses. Centrée sur le problème de la coordination et ses conséquences, cette approche invite à considérer les pratiques scientifiques comme une succession d'itérations épistémiques et à faire du dynamisme inhérent à cette succession d'itérations l'objet d'étude premier de la philosophie des sciences. Les itérations sont des ensembles de pratiques identifiées de façon pragmatique et à différentes échelles dans le but de cerner des discours dont la comparaison pourrait se révéler informative. Ainsi, des itérations de grande échelle, comme la phylogénétique dans son ensemble en tant qu'itération de la systématique, peuvent être comparées. Un regard plus précis peut aussi être adopté en cernant, par exemple, des itérations successives au sein de la phylogénétique, comme l'itération traditionnelle centrée sur les modèles arborescents et l'itération récente centrée sur les modèles réseaux. Dans tous les cas, le cohérentisme historique invite à comparer des itérations et à formuler des explications permettant de comprendre les transformations subies par les pratiques scientifiques.

Si la confrontation entre hypothèses, modèles et théories, d'un côté, et phénomènes, de l'autre, est indéniablement un déterminant important du dynamisme scientifique, il est loin d'être le seul. Le principe de respect (ou le poids de la tradition), l'injonction au progrès, l'importation d'outils d'autres disciplines, les développements matériels et mathématiques, les contingences sociohistoriques et la structure sociale des communautés scientifiques sont aussi d'importants déterminants de ce dynamisme. En tant que philosophe, je me suis concentré, dans cette thèse, sur les déterminants dont les traces peuvent être observées par l'analyse qualitative d'éléments du discours scientifique et tout particulièrement ceux qui apparaissent au sein

d'articles et de livres publiés après évaluation par les pairs. En revanche, il convient de noter que le cohérentisme historique incite au développement de pratiques pluridisciplinaires afin de considérer un maximum de facteurs.

La perspective synchronique, dont l'importance a été soulignée par van Fraassen (2008), illustre bien l'injonction à considérer un maximum de facteurs pour comprendre les dynamiques des pratiques scientifiques. Cette perspective suggère de croiser regards anhistoriques et historiques afin de réaliser des analyses épistémologiques. Il s'agit d'adopter la logique d'une théorie stabilisée pour interpréter les inférences du champ disciplinaire selon son propre point de vue tout en croisant ce regard avec une perspective historique qui met en lumière l'assemblage graduel des systèmes de pensée qui caractérisent les cultures scientifiques. J'ai appliqué cette perspective synchronique à l'étude des pratiques phylogénétiques, tout en intégrant à la méthode d'analyse des principes issus d'autres textes associés au cohérentisme historique. En ce sens, j'ai préconisé une focalisation du propos sur les modèles et leur inscription dans la pratique (Mari 2005; Morrison 1998; Tal 2016) et j'ai accordé une grande importance aux négociations entre le principe de respect et l'injonction au progrès (Chang 2004) qui forment les successions d'itérations épistémiques. L'ensemble des outils que j'ai extraits des textes associés au cohérentisme historique sont compilés dans la conclusion du chapitre 2.

Le cohérentisme historique, comme reconstruit au chapitre 2, est donc le cadre théorique que j'ai utilisé pour analyser les pratiques phylogénétiques dans la deuxième moitié de la thèse (chapitres 5 à 7). La nécessité de développer ce cadre théorique pourrait cependant être contestée sur la base du fait que la philosophie des sciences offre déjà des outils d'analyse permettant d'appréhender les pratiques classificatoires, à savoir ceux inscrits au sein de la tradition des espèces naturelles. Aux chapitres 3 et 4, j'ai donc présenté cette tradition et argumenté contre son utilisation pour l'analyse épistémologique des pratiques classificatoires en sciences, justifiant par le fait même l'importation du cohérentisme historique de la philosophie de la mesure vers la philosophie de la classification.

Au chapitre 3, j'ai exploré le corpus de textes de la tradition des espèces naturelles qui s'inscrit en philosophie des sciences. L'idée au cœur de cette tradition est que certaines

catégories, à identifier au sein de celles que les pratiques scientifiques « découvrent » et mobilisent, sont des espèces naturelles, c'est-à-dire des entités avec un statut métaphysique ou épistémique privilégié. En philosophie des sciences, ce constat est doublé de la croyance selon laquelle l'analyse épistémologique des pratiques classificatoires devrait être centrée sur la notion d'espèce naturelle. Face à l'impressionnante étendue du corpus de textes concernés, j'ai opté pour une approche descriptive qui cible des textes représentatifs des principaux courants au sein de la tradition.

Les principales approches des espèces naturelles sont l'essentialisme (Bird 2009; Devitt 2008; LaPorte 2004), le pluralisme (Dupré 1993; Boyd 1991, 1999a) et le monisme non essentialiste (Khalidi 2013; Slater 2015). L'essentialisme stipule que les espèces naturelles sont des regroupements d'entités qui partagent une essence, celle-ci étant habituellement considérée comme une microstructure inhérente aux entités (l'ADN, par exemple) qui expliquerait la majorité de leurs traits (et donc la similarité entre les membres d'une espèce naturelle). Les conceptions pluralistes de Boyd et Dupré ont été développées pour souligner que l'essentialisme est incapable de rendre compte de l'ensemble des catégories qui devraient être considérées comme des espèces naturelles. Dupré, par son réalisme dévergondé (*promiscuous realism*), a suggéré que toute catégorie épistémiquement utile devrait être considérée comme une espèce naturelle et que la multiplicité des découpages du monde qui résulte de ce pluralisme représente différents découpages *réels*. Boyd, en plus de partager plusieurs intuitions avec Dupré, propose une description détaillée d'espèces naturelles qui ne se conforment pas à l'essentialisme, soit les groupes homéostatiques de propriétés (GHP). Selon cette théorie, les membres d'une espèce naturelle GHP sont des ensembles de propriétés qui tendent à exister ensemble puisqu'un mécanisme causal (homéostatique) assure leur cooccurrence.

Depuis la formulation de ces deux théories pluralistes (le réalisme dévergondé et la théorie GHP inscrite dans le pluralisme de Boyd), un grand nombre de publications concernant les espèces naturelles ont ressuscité l'espoir de mettre à jour une théorie moniste. Toujours au troisième chapitre, j'ai présenté deux de ces théories qui sont fortement opposées. La théorie de Khalidi, comme l'essentialisme, met la description ontologique des espèces naturelles à l'avant-plan. Plus précisément, il décrit les espèces naturelles comme étant des nœuds au sein de réseaux

de causalité. Ce serait donc la nature des rapports causaux internes et externes des occurrences qui permettrait de les identifier comme les membres d'une espèce naturelle. Slater, au contraire, met l'accent sur le statut épistémique des espèces naturelles, tout comme Boyd et Dupré. En effet, le fait que certaines catégories soient aptes à soutenir les inférences des différentes disciplines scientifiques est suffisant, selon lui, pour leur attribuer le statut métaphysique d'espèce naturelle.

J'ai ainsi brossé un portrait de la tradition des espèces naturelles en illustrant la diversité des approches disponibles. À la suite de Hacking (1991), j'ai argumenté que la tradition des espèces naturelles est bien unitaire malgré cette hétérogénéité. Plus précisément, j'ai défendu que même des approches aussi différentes que celles de Khalidi et de Slater sont structurées sur la base d'un modèle partagé, que j'ai appelé l'arche des espèces naturelles (Figure 3.1). Cette arche, qui s'est cristallisée dès le 19^e siècle, est composée de quatre éléments architecturaux : deux piliers, une esse de soutien et un arc. Les piliers présentent chacun une croyance, soit celle selon laquelle des catégories détiennent un statut métaphysique particulier et celle selon laquelle des catégories détiennent un statut épistémique particulier. L'esse de soutien représente le lien logique entre ces deux piliers, un lien qui solidifie l'ensemble de la structure. Ce lien peut prendre deux formes : les particularités métaphysiques des catégories (comme leur essence) expliquent leur statut épistémique privilégié (comme dans l'approche de Khalidi), ou à l'inverse, le statut métaphysique est inféré sur la base de l'utilité épistémique des espèces naturelles (comme dans l'approche de Slater). L'arc de l'arche, qui repose sur les deux piliers ainsi solidifiés, est la croyance selon laquelle il est important de se doter d'une théorie des espèces naturelles afin de comprendre épistémologiquement les pratiques classificatoires. Toute théorie considérant que l'étude des pratiques classificatoires requiert le développement de théories portant sur des catégories épistémiquement ou ontologiquement privilégiées s'inscrit ainsi dans la tradition des espèces naturelles en adhérant au modèle tout juste décrit.

Dans le chapitre 4, j'ai défendu une posture éliminativiste à l'égard des espèces naturelles. Après avoir décrit deux arguments éliminativistes déjà présents au sein du corpus de textes étudiés, soit celui de Ludwig (2018) fortement inspiré de celui de Hacking (2007a) et celui de Brigandt (2020), j'ai formulé un argument qui se veut complémentaire aux deux autres. Ludwig

affirme que la pluralité des théories des espèces naturelles et l'utilité de l'ensemble de ces théories suggèrent que le concept général d'espèce naturelle devrait être abandonné au profit de concepts plus précis. Il s'agit donc d'un argument de l'ordre du pluralisme éliminativiste (Ereshefsky 1992) qui maintient les catégories scientifiques comme point focal des analyses épistémologiques. L'argument de Brigandt (2020) vise quant à lui à rejeter la naturalité des catégories scientifiques. En plus de pouvoir engendrer des situations sociales problématiques, cette naturalité, nous dit Brigandt, occulte le fait que les catégories scientifiques sont forgées par un ensemble des valeurs (épistémiques ou non). Il suggère donc de remplacer l'étude des espèces naturelles par l'étude de la contribution des catégories scientifiques à la réalisation d'objectifs précis ou généraux, ce qui permet de tenir compte activement de l'impact des valeurs en sciences (un constat qui s'inscrit tout naturellement au sein du cohérentisme historique).

Afin de détailler mon argument éliminativiste à l'égard des espèces naturelles, j'ai amorcé un dialogue avec un plus vieux texte de Brigandt (2009) au sein duquel il défend l'utilité épistémologique du constat selon lequel l'homologie est une espèce naturelle GHP. Selon Brigandt (2009), ce constat est utile pour au moins deux raisons. D'abord, il permet d'identifier un fait métaphysique qui explique l'utilité épistémique de l'homologie (une explication épistémologiquement fort intéressante). Ce constat est toutefois inexact puisque l'étiquette d'espèce naturelle est associée à des traits biologiques seulement une fois que ces traits ont été identifiés comme des homologues. Or, l'identification d'homologues, en biologie de l'évolution, implique qu'on ait déjà expliqué pourquoi il y a similitude entre les traits biologiques concernés : la similarité serait le fruit d'une ascendance commune. C'est donc bel et bien la généalogie des organismes concernés qui explique (dans ce contexte) la similarité et non le fait que les homologues forment une espèce naturelle. Au mieux, la théorie GHP vient seulement réitérer une information déjà présente au sein des pratiques scientifiques et est donc peu apte à guider l'analyse épistémologique.

Ensuite, Brigandt argumente qu'il est nécessaire d'incarner la connaissance au sujet de liens de similarité entre les membres d'une catégorie par l'entremise d'un concept associé à une espèce naturelle. Au contraire, je considère que les concepts sont certainement utiles pour simplifier la référence aux phénomènes, mais cela n'implique aucunement qu'il s'agisse de

concepts associés à des espèces naturelles. Par exemple, la catégorie « résistance aux antimicrobiens » peut être utilisée pour référer à un ensemble de phénomènes, de sorte à nous informer sur les phénomènes ainsi désignés sans que cette catégorie n'ait à former une espèce naturelle (la résistance aux antimicrobiens étant un phénomène extrêmement hétérogène). Étiqueter ce genre de catégories comme une espèce naturelle rendrait le concept trivial, même si cette catégorie incarne bel et bien des connaissances par rapport à certains phénomènes.

À ces deux réponses au texte de Brigandt (2009), j'ai ajouté deux arguments qui remettent en question l'utilité épistémologique des espèces naturelles. Premièrement, les théories des espèces naturelles sont plus adéquatement décrites comme des mécanismes de légitimation des pratiques classificatoires en sciences que comme des théories pour l'analyse épistémologique de ces pratiques. En effet, en acceptant de fait que certaines catégories scientifiques aient des statuts métaphysiques ou épistémiques particuliers (les piliers de l'arche), ces théories ne font qu'offrir un discours philosophique qui renforce la validité des pratiques qui mobilisent ces catégories.

Deuxièmement, comme le cohérentisme historique le souligne à merveille, les pratiques scientifiques sont le fruit de réseaux de croyances complexes qui structurent les itérations successives et qui permettent notamment de rendre observable ce qui ne l'est pas directement. Pour bien comprendre les pratiques classificatoires en sciences, il faut se pencher sur de tels systèmes dans leur ensemble plutôt que de se concentrer simplement sur les catégories qui en sont le produit. La phylogénétique est un bon exemple des apories de la tradition des espèces naturelles : la reconstruction de systèmes phylogénétiques, soit l'entreprise principale de la discipline, reste absente des analyses épistémologiques centrées sur les catégories scientifiques mises en relation au sein de tels systèmes.

Mon argument éliminativiste, bien que complémentaire à ceux de Brigandt et de Ludwig, se veut donc plus fort. Ces deux auteurs considèrent en effet que les catégories scientifiques, une fois dénudées de leur naturalité et dissociées du monisme inhérent au concept d'espèce naturelle, sont des points focaux adéquats pour l'analyse des pratiques classificatoires en sciences. En contraste, je suggère que l'étude des pratiques classificatoires devrait être centrée

sur les agencements de croyances et les différents facteurs qui déterminent ces agencements plutôt que sur les catégories qui n'en sont qu'un des produits.

C'est ainsi que les quatre premiers chapitres de la thèse m'ont permis de défendre une approche épistémologique bien précise à l'égard des pratiques classificatoires : celles-ci sont composées d'itérations épistémiques successives, chacune d'entre elles ayant à négocier avec les difficultés liées à la coordination. La philosophie de la classification que j'ai proposée dans cette thèse est une boîte à outils permettant l'analyse de cette succession d'itérations tout en gardant à l'avant-plan la charge théorique des observations. Afin de démontrer l'efficacité de cette approche, et du même coup, de contribuer à certains débats importants en philosophie de la biologie, les trois derniers chapitres de la thèse mobilisent le cohérentisme historique pour analyser des pratiques phylogénétiques contemporaines. Je me suis particulièrement intéressé au contraste entre les modèles arborescents, qui sont au cœur de la phylogénétique traditionnelle, et les modèles réseaux, qui sont considérés par beaucoup comme ayant engendré une rupture au sein de la discipline.

Au chapitre 5, j'ai présenté un développement important de la biologie qui a drastiquement transformé la phylogénétique, soit l'accès grandissant aux données moléculaires qui caractérise la deuxième moitié du 20^e siècle (O'Malley 2016; Sapp 2009). Considérant la centralité des données moléculaires au sein de la phylogénétique contemporaine, il est nécessaire de saisir les changements subis par cette discipline à la suite de l'introduction de ce type de données. J'ai centré mon propos sur la notion d'homologie pour deux raisons. Premièrement, ce concept traduit efficacement la structure épistémique de la phylogénétique et est ainsi un excellent témoin de l'évolution de la discipline. Deuxièmement, cela m'a permis de souligner l'efficacité analytique du cohérentisme historique en contraste avec la tradition des espèces naturelles, l'homologie étant en effet souvent considérée comme une espèce naturelle.

Que l'analyse phylogénétique porte sur des données moléculaires ou des données morphologiques, l'homologie se définit comme un lien de similarité entre des traits issu de l'ascendance commune des entités qui portent ces traits (Inkpen et Doolittle 2016; Kapli, Yang et Telford 2020; Rieppel 2015). De plus, son rôle au sein de la structure inférentielle demeure le

même. Dans les deux cas, l'homologie incarne ce lien similarité-généalogie qui est nécessaire à l'observation (indirecte) des liens phylogénétiques entre unités taxonomiques. Or, sous ces éléments de continuité se cache une distinction importante. Les modèles pour l'analyse de la similarité mobilisés pour les données génétiques requièrent des hypothèses relatives aux transformations subies par les traits, soit des séquences génétiques (Delsuc, Brinkmann et Philippe 2005; Morrison, Morgan et Kelchner 2015). Au contraire, les critères pour l'analyse de la similarité morphologique ne mobilisent aucune hypothèse de ce genre (Remane 1952; Patterson 1982).

C'est donc dire qu'une importante distinction existe quant à l'opérationnalisation du concept d'homologie, une distinction qu'une analyse épistémologique centrée sur la théorie ou la définition du concept d'homologie pourrait manquer. Au contraire, en s'intéressant aux modalités qui permettent la coordination entre la théorie darwinienne de l'évolution (comprise comme allant de pair avec la notion de lignée) et les phénomènes, le cohérentisme historique invite à explorer les modèles qui rendent possible cette coordination. L'inscription d'hypothèses transformistes au sein de ces modèles, dans le cas de la phylogénétique pratiquée avec des données moléculaires, est tributaire de l'importance accordée à la transformation des lignées qui en vient à profondément déterminer notre compréhension de ce qu'est la similarité biologique. Ce constat renforce de fait la validité du modèle hélicoïdal pour rendre compte des dynamiques scientifiques : les analyses de similarité, centrées sur l'homologie, ont contribué à faire émerger la théorie darwinienne de l'évolution dont la focalisation sur la formation de lignées divergentes vient transformer notre regard sur la similarité.

Au chapitre 6, j'ai poursuivi l'analyse de l'hélice qui caractérise le système épistémique de la phylogénétique. J'y ai analysé les pratiques de la phylogénétique traditionnelle, soit l'itération phylogénétique centrée sur les modèles arborescents qui domine la systématique depuis les années 1950 (avec ou sans molécules). Pour ce faire, j'ai présenté ce qui a été appelé la pensée arborescente (O'Hara 1997), un cadre théorique abstrait qui structure la phylogénétique traditionnelle. Mon analyse a été réalisée à trois niveaux d'abstraction distincts. D'abord, je me suis penché sur un texte à teneur épistémologique qui analyse les tendances dominantes de la phylogénétique traditionnelle et qui décrit la pensée arborescente (O'Hara 1997). L'auteur du

texte, O'Hara, y met en exergue ce qu'il considère être les principes fondamentaux de cette pensée arborescente, soit de considérer les clades comme des individus historiques liés (phylogénétiquement) les uns aux autres (et non des occurrences indépendantes d'un rang taxonomique comme l'espèce ou le genre) et le rejet de l'association entre évolution et complexification.

Après avoir présenté la pensée populationnelle à ce niveau d'abstraction, j'ai offert un peu plus de concrétude en analysant l'inscription de cette pensée au sein d'un programme de recherche, soit la phylogénomique décrite par Delsuc et ses collègues (2005). Cette analyse m'a permis de présenter l'influence que le tournant génomique a eue sur la phylogénétique. Sans bouleverser significativement la structure inférentielle de la discipline, l'accès aux données massives issues de l'amélioration des techniques de séquençage et l'augmentation de la puissance calculatoire disponible ont changé l'échelle de la phylogénétique. Plutôt que de limiter les analyses à un nombre restreint de gènes, il est devenu possible de réaliser des analyses de similarité plus englobantes par l'utilisation de nombreux gènes ou même de données concernant directement le génome. Or, le texte programmatique de Delsuc et ses collègues démontre que cette transition portant sur la quantité de données traitées peut être effectuée en harmonie avec la pensée arborescente. En effet, l'adhésion des auteurs à ce cadre de pensée ressort indéniablement de l'analyse épistémologique de leur discours. Il convient cependant de préciser que le lien entre phylogénomique et pensée arborescente est loin d'être un lien de nécessité, comme suggéré par le texte Delsuc et ses collaborateurs. En effet, les données phylogénomiques ont par ailleurs été la source de remises en question de plusieurs éléments de ce cadre de pensée (Charlebois, Beiko et Ragan 2003).

L'étude de ces deux textes permet de réaliser une analyse critique de la phylogénétique traditionnelle. En croisant la logique propre à la pensée arborescente (qui prend pour acquises diverses associations d'idées, notamment celles qui rendent possible la coordination entre théorie et empirie) avec une perspective diachronique, j'ai mis en lumière trois caractéristiques importantes de la phylogénétique centrée sur les modèles arborescents.

Premièrement, cette itération de la systématique est structurée par une série d'amalgames. J'ai utilisé le terme d'amalgame pour référer à des assemblages d'idées qui sont présentés comme étant des unions nécessaires alors que les idées les composant sont tout à fait indépendantes. L'identification de tels amalgames est indéniablement importante dans le cadre de l'épistémologie qui cherche notamment à décrire de façon précise les entreprises de construction de savoir. Je considère aussi qu'il est important de décortiquer les amalgames présents au sein d'une discipline scientifique afin de favoriser les débats constructifs au sein de cette discipline. Le premier amalgame consiste à lier le modèle arborescent avec le rejet de la pensée du groupe et de la pensée du développement. S'il est vrai que les pratiques contemporaines rejettent ces deux perspectives sur les phénomènes évolutifs, il convient tout de même de souligner que le modèle arborescent a été utilisé, par le passé, en lien avec ces deux pensées.

Le deuxième amalgame consiste à lier de façon nécessaire le modèle arborescent et la phylogénétique. Cette association d'idées est évidemment inexacte. La présentation de la phylogénétique centrée sur les réseaux, au chapitre 7, en est le meilleur témoignage. Cet amalgame est issu de la place centrale que le principe de divergence (en vertu duquel Darwin a postulé que les relations phylogénétiques entre organismes correspondent, ont correspondu et correspondront au modèle arborescent) occupe en phylogénétique. Le troisième amalgame revient à considérer que l'étude de l'évolution par voie de sélection naturelle requiert l'étude des dynamiques de formations de lignées. J'ai démontré ailleurs (Papale 2021) que l'évolution par voie de sélection naturelle peut se produire même en absence d'entités biologiques qui forment des lignées. J'ai exploré la portée philosophique de ce constat au chapitre 7 puisque la phylogénétique centrée sur les réseaux véhicule aussi ce troisième amalgame.

La seconde caractéristique de la phylogénétique traditionnelle que j'ai mise en lumière est la présence d'une posture moniste et réaliste au sein de cet ensemble de pratiques. En effet, un principe ontologique (soit une croyance structurante des systèmes épistémiques qui est injustifiée, voire injustifiable) est présent au sein de la phylogénétique traditionnelle. Il s'agit du principe d'histoire évolutive unique selon lequel il n'y aurait qu'un système phylogénétique à reconstruire. Ainsi, toutes les données devraient converger sur ce système unique. Doublé du

modèle arborescent, le principe d'histoire évolutive unique stipule que l'Arbre du vivant, au singulier, est plus qu'un modèle. Il s'agit d'une représentation de l'histoire réelle des unités taxonomiques.

Le fait même qu'une posture de la sorte existe au sein de la phylogénétique traditionnelle souligne l'importance de certains courants philosophiques pour le regard des praticiens et des praticiennes sur leur propre entreprise de construction de connaissances. En effet, le réalisme et le monisme inhérents à la phylogénétique traditionnelle témoignent d'une posture fondationnaliste : les phénomènes et les données qui en sont issues seraient les déterminants privilégiés des dynamiques scientifiques; ils seraient les fondations de la connaissance. Ce constat entre en conflit avec la présence du problème de la coordination au sein de la structure épistémique de la phylogénétique : notre accès aux phénomènes dépend d'éléments théoriques préalablement acceptés (comme le modèle arborescent ou le lien similarité-phylogénie) sur la base d'itérations épistémiques passées, de sorte que les données ne peuvent être des fondations absolues; les données elles-mêmes reposent minimalement sur le passé de la discipline et la légitimité que celle-ci a acquise au travers du temps.

Mon analyse qui mène à ces trois constats critiques a d'abord été basée sur les textes de O'Hara et de Delsuc et ses collègues, des textes qui sont à un niveau d'abstraction relativement élevé et qui datent. Dans la deuxième moitié du chapitre, j'ai donc analysé un débat récent, un dialogue entre deux groupes de chercheurs et de chercheuses ayant été publié notamment dans la prestigieuse revue *Nature Microbiology* (Aouad et al. 2019; Sorokin et al. 2017, 2019). Ce débat, au premier regard et selon les chercheurs et chercheuses impliqués, porte sur le placement phylogénétique d'un clade d'archées récemment découvert, soit *Methanonatronarchaeia*. Or, une analyse épistémologique du débat témoigne qu'il s'agit en fait d'un débat méthodologique portant sur les modalités des procédures permettant de rendre les liens phylogénétiques observables (un débat qui a des conséquences notamment sur le placement de *Methanonatronarchaeia*). Ce débat m'a permis de démontrer que les trois constats critiques précédents sont bien d'actualité au tournant de la troisième décennie de notre siècle. Le troisième constat est particulièrement bien illustré au sein du débat : alors même que les articles traitent des modalités des procédures phylogénétiques qui permettent le traitement des

données, les auteurs et autrices espèrent que l'accumulation d'une plus grande quantité de données pourra régler la question.

La pensée arborescente ayant été ainsi analysée en détail, il est devenu possible de comparer la phylogénétique traditionnelle avec la phylogénétique centrée sur les modèles réseaux. Pour ce faire, le chapitre 7 a présenté cette plus récente itération de la phylogénétique. La notion de réseau, telle qu'elle est utilisée en phylogénétique, réfère à des systèmes d'affinités entre unités taxonomiques qui, contrairement au modèle arborescent, présentent de la réticulation, des liens transversaux entre lignées (Huson, Rupp et Scornavacca 2010; Morrison 2014a, 2014b). Le modèle arborescent peut être décrit en adoptant le vocabulaire de la théorie des graphes comme étant un réseau acyclique au sein duquel les nœuds intérieurs ont des degrés maximaux de trois (un maximum d'un ancêtre et de deux lignées filles). Le degré des nœuds terminaux (les taxons n'ayant pas de descendance) est quant à lui limité à un. En ce sens, le modèle arborescent est dominé par la présence de bifurcations (une lignée fait place à deux lignées). Au contraire, les modèles réseaux relâchent la contrainte associée à la bifurcation et donc à la divergence. Le degré des nœuds peut être plus grand que trois, et lorsque les graphes sont dirigés (c'est-à-dire que les arêtes liant les nœuds ont une direction), ce relâchement concerne spécifiquement les degrés entrants (le nombre d'arêtes menant au nœud). En effet, la particularité des réseaux phylogénétiques est de permettre et de modéliser la pluralité des sources d'une lignée phylogénétique. Dans les faits, le nombre de lignées ancestrales auxquelles un nœud peut être lié est souvent limité à deux; l'utilisation de modèles réseaux en phylogénétique témoigne tout de même d'une ouverture logique aux dynamiques évolutives menant à la réticulation des liens de parenté (transferts génétiques latéraux, hybridation, endosymbiose, etc.).

Pour bien saisir la teneur de ce dernier constat, il faut revenir sur la distinction entre réseaux abstraits et explicites (une distinction qui s'applique aussi aux modèles arborescents). Les réseaux abstraits sont des représentations des affinités entre unités taxonomiques qui cherchent à rendre explicite le fait que ces affinités sont pluridirectionnelles. Ces réseaux sont principalement utilisés pour représenter des données qui sont considérées, dans une perspective arborescente, comme étant conflictuelles. L'idée de conflit est ici entièrement soumise au patron

arborescent : les données sont en conflit en ce sens qu'elles soutiennent des topologies arborescentes différentes. En acceptant les liens transversaux entre lignées (ce qui engendre des nœuds de degrés supérieurs à trois), les réseaux abstraits permettent de représenter la pluralité des topologies, les conflits qui les opposent et ainsi de représenter de façon plus complète le jeu de données.

Or, ces représentations réticulées peuvent aussi être interprétées comme des systèmes phylogénétiques à part entière. Les réseaux abstraits deviennent alors explicites. Un réseau explicite, en phylogénétique, est un réseau qui représente une hypothèse par rapport aux dynamiques de formation de lignées. Ainsi, ce réseau doit être dirigé et avoir une racine, un point de départ, un ancêtre commun à l'ensemble des taxons étudiés. La réticulation, qui, au sein des réseaux abstraits, n'est qu'une représentation de conflits entre données, devient ainsi un phénomène biologique concret, soit la mise en commun de deux lignées pour en former une troisième. Ce phénomène est extrêmement hétérogène puisque des processus aussi distincts que l'hybridation des espèces, le transfert de matériel génétique entre génomes ou l'endosymbiose engendrent de la réticulation au sein des dynamiques de formation de lignées et détiennent tous une signature phylogénétique qui leur est propre.

Cette description des modèles réseaux et de leur utilisation en phylogénétique permet de comparer leur utilisation à celle des modèles arborescents. J'ai réalisé cette analyse de deux façons distinctes. D'abord, j'ai brossé le portrait des affinités entre les deux itérations de la phylogénétique, ce qui m'a permis d'insister à la fois sur les liens de similarité et sur les différences marquantes. En bref (les résultats de l'analyse sont synthétisés dans le tableau 7.1, les deux modèles structurent des façons de pratiquer la systématique qui sont largement similaires : l'impératif classificatoire, la focalisation sur les lignées et l'utilisation des mêmes modèles d'analyse de la similarité pour inférer la généalogie sont les meilleurs témoins de cette structure partagée. Nonobstant cette continuité indéniable, des différences marquées doivent être soulignées. L'ouverture logique aux liens généalogiques transversaux inscrite au cœur de la phylogénétique mobilisant les modèles réseaux représente une rupture qui peut être interprétée comme un progrès : alors que la phylogénétique traditionnelle considère certaines données

comme du bruit, l'approche réseau permet d'utiliser cette information pour inférer la présence de phénomènes dont l'importance théorique est de plus en plus reconnue.

Une fois ce tableau comparatif établi, je l'ai utilisé dans la deuxième moitié du chapitre 7 pour confronter les deux itérations épistémiques, soit celle centrée sur les réseaux et celle centrée sur les arbres, à cinq débats qui gagnent à être informés par une analyse épistémologique rigoureuse des pratiques phylogénétiques. Les trois premiers débats, observés au sein du corpus disciplinaire à l'étude, concernent 1- la compatibilité des modèles réseaux avec le cadre théorique (néo-)darwinien, 2- la teneur réaliste des modèles considérés et 3- l'utilité respective des deux modèles. À la suite d'autres philosophes et de biologistes (Baptiste 2007; Charlebois, Beiko et Ragan 2003), j'ai démontré que les modèles réseaux sont tout à fait compatibles avec la théorie de l'évolution par voie de sélection naturelle et qu'ils renforcent de surcroît la remise en question du principe de divergence au sein de cette théorie, une remise en question constructive. De plus, ce premier débat permet de souligner une force du cohérentisme historique, soit de décomposer des schèmes de pensées en croyances constitutives.

En ce qui concerne le deuxième débat, je l'ai utilisé pour réitérer ma posture à l'égard du réalisme et de l'antiréalisme. Ces postures philosophiques, lorsqu'elles sont utilisées pour traiter de modèles scientifiques, concernent la conformité de ces modèles par rapport aux phénomènes tels qu'ils existeraient indépendamment de nous. De telles postures s'effritent face à la charge théorique des observations. En effet, le problème de la coordination décrit avec précision les contraintes liées à tout contact épistémique avec les phénomènes. En conséquence, les postures réalistes et antiréalistes à l'égard de certains modèles traitent, par définition, d'un aspect des phénomènes auquel nous n'avons point accès, soit leur existence indépendante de nos schèmes de pensée.

De ce fait, la réalité des modèles est un pauvre critère pour leur évaluation épistémique (ou épistémologique). Au contraire, leur utilité, comprise au sens des défis propres aux structures épistémiques qui les mobilisent, est un critère qui permet des débats constructifs, ce qui nous amène directement au troisième débat considéré. En évaluant l'utilité des modèles réseaux et arborescents en phylogénétique, un constat s'impose. Les deux types de réseaux sont utilisés,

pour l'instant, de façon complémentaire. La puissance calculatoire nécessaire à la réalisation d'inférences de réseaux phylogénétiques étant immense, les modèles arborescents demeurent utiles si ce n'est qu'en facilitant le traitement de plus grandes quantités de données. De surcroît, un nombre important de procédures menant à la reconstruction de réseaux phylogénétiques utilisent le modèle arborescent. Ainsi, il s'agit simplement d'accorder les procédures aux défis que l'on espère relever et de rester sobre face aux déclarations selon lesquelles un type de modèle (arborescent ou réseau) serait univoquement supérieur à l'autre. De telles simplifications découlent d'analyses épistémologiques imprécises.

Le chapitre 7 et le corps de la thèse se concluent sur mon identification de deux débats absents au sein du corpus de textes étudié, mais qui méritent l'attention des biologistes et philosophes. Ces deux débats sont ancrés dans les points communs liant les itérations phylogénétiques centrées sur les réseaux et les arbres. Le premier relève de la focalisation de la phylogénétique sur les dynamiques de formation de lignées. Le second traite de l'impératif classificatoire.

Comme évoqué précédemment, les deux itérations épistémiques à l'étude sont liées par leur focalisation sur l'étude des lignées. Or, j'ai démontré ailleurs (Papale 2021) que la théorie de l'évolution par voie de sélection naturelle permet d'expliquer le changement (ou la stabilité) au sein de populations d'entités qui forment des lignées, certes, mais également au sein de population d'entités qui sont indépendantes de ces lignées (voir aussi Charbonneau 2014 et Roughgarden 2020). J'ai aussi démontré au chapitre premier de cette thèse que la phylogénétique vise notamment à fournir un contenu empirique à la théorie (néo-)darwinienne de l'évolution. La mise en relation de ces deux constats témoigne d'un décalage au sein de l'hélice épistémique qui caractérise la biologie de l'évolution : les démarches empiriques liées à la théorie de l'évolution ne traitent que d'une partie des phénomènes que celle-ci permet d'expliquer. Nombre de phénomènes d'évolution, comme celui impliquant la résistance aux carbapénèmes, le gène *bla_{NDM}* et *Acinetobacter baumannii*, ne peuvent simplement pas être réduits efficacement à l'histoire des lignées impliquées. Ainsi, une question se doit d'être posée : le rôle central assigné à la phylogénétique au sein de la biologie de l'évolution est-il justifié? Si le lecteur et la lectrice peuvent certainement se douter de la posture que j'adopterais face à une telle question, l'objectif

de ce segment de la thèse était simplement de démontrer l'importance d'amorcer un débat sur le sujet.

Dans une optique similaire, le cinquième débat concerne l'impératif classificatoire présent en biologie. Comme l'omniprésence de la nomenclature linnéenne en témoigne, les descriptions classificatoires et phylogénétiques des entités biologiques occupent une place importante au sein de la discipline. J'ai donné de bonnes raisons de croire qu'il pourrait être édifiant de prioriser d'autres types de descriptions du monde biologique en insistant sur les descriptions quantitatives. Par définition, les descriptions classificatoires occultent l'information qui réside dans les différences entre les individus puisqu'elles décrivent ces individus sur la base de leurs points communs qui justifient leur appartenance à une catégorie donnée. Particulièrement dans un contexte darwinien, cette situation est problématique puisque les différences de fitness entre individus expliquent comment la sélection naturelle peut en venir à transformer les populations dans une direction donnée. Des descriptions du monde vivant qui négligent ce point semblent donc peu appropriées à l'étude de l'évolution. Ainsi, je suggère de mettre l'accent sur la quantification des affinités (morphologiques, génétiques, généalogiques, fonctionnelles, topologiques, etc.) entre entités biologiques pour tenir compte des différences entre individus. Cette suggestion s'inscrit en cohérence avec des procédures présentes au sein de la phylogénétique, mais qui tendent à être soumises à l'injonction classificatoire.

En résumé, le cohérentisme historique est une approche épistémologique qui met l'accent sur les analyses comparatives de systèmes épistémiques. En tant que philosophe, je me suis concentré sur les différentes croyances qui composent ces systèmes et leur implantation au sein de procédures d'observations concrètes. Cela m'a permis de développer une philosophie de la classification qui doit être considérée comme une boîte à outils pour l'analyse des pratiques classificatoires plutôt que comme un discours sur le statut métaphysique des catégories scientifiques ou sur ce qu'est une bonne classification. La philosophie mise de l'avant insiste sur le besoin de faire des analyses localisées et sur la réflexivité des pratiques scientifiques qui doit trouver un écho dans la réflexivité des pratiques épistémologiques. Les contributions de la thèse peuvent être divisées en deux groupes. Le premier concerne la philosophie des sciences et l'épistémologie plus généralement :

- Le problème de la coordination s'incarne dans tout ensemble de pratiques qui implique l'ancrage empirique d'éléments symboliques; il s'incarne donc au sein des pratiques classificatoires et de la structure épistémique de la phylogénétique.
- La structure épistémique des pratiques scientifiques a une dimension diachronique; chaque itération épistémique s'appuie sur des itérations passées.
- Le cohérentisme historique est une approche adéquate pour l'analyse comparative des itérations épistémiques des disciplines classificatoires.
- La tradition des espèces naturelles, au contraire, est lacunaire.

Le second ensemble de contributions de cette thèse relève plutôt de la philosophie de la biologie :

- Les pratiques phylogénétiques contemporaines sont le fruit d'un processus historique qui a engendré des amalgames de croyances logiquement et historiquement indépendantes.
- La correspondance entre modèles ou théories, d'un côté, et phénomènes, de l'autre, n'est qu'un élément parmi tous ceux qui confèrent à la phylogénétique sa légitimité épistémique; le poids de la tradition (principe de respect) est un de ces éléments et il occupe une place importante.
- Les phylogénéticiens et phylogénéticiennes tendent à adopter des postures philosophiques (monisme, réalisme et fondationnalisme) incompatibles avec les conséquences du problème de la coordination.
- Les modèles réseaux contribuent à opérer une ouverture logique au sein de la biologie de l'évolution, soit l'ouverture aux phénomènes qui dépassent la formation de lignées verticales.
- Nonobstant ce progrès, l'ensemble de la structure inférentielle et des procédures concrètes de la phylogénétique centrée sur les réseaux est le même que celui de la phylogénétique traditionnelle.
- La focalisation de la biologie de l'évolution sur les dynamiques de formation de lignées doit être réévaluée.

- La dimension historique de la phylogénétique est soumise à sa dimension classificatoire.
- La priorisation de méthodes descriptives classificatoires, en biologie, doit être réévaluée; des méthodes quantitatives pourraient prendre le relais.

C'est donc dire que cette thèse contribue à l'amélioration des analyses épistémologiques des pratiques classificatoires. L'efficacité du cohérentisme historique a été démontrée par son utilisation dans le cadre d'une analyse de pratiques phylogénétiques contemporaines. L'analyse qui en résulte aurait été impossible dans une perspective anhistorique qui occulte la charge théorique des observations. Cette charge théorique doit être à l'avant-plan de l'analyse de tout discours gnoséologique puisque sans elle, il est impossible de fournir un ancrage empirique aux différents modèles, croyances et théories qui traversent ces discours.

La recherche effectuée dans le cadre de ce projet doctoral ouvre plusieurs opportunités d'exploration future. D'un point de vue biologique, certaines propositions faites au chapitre 7 méritent d'être approfondies, notamment pour déterminer si un soutien empirique peut leur être fourni. Par exemple, j'ai suggéré que l'étude de l'évolution gagnerait à être découplée de l'étude des lignées. Si la modélisation de dynamiques au sein desquelles la formation de lignées est négligeable est déjà sur la table (Roughgarden 2020), il serait intéressant d'identifier des populations d'entités biologiques inédites qui correspondent aux conditions de possibilité de l'évolution par voie de sélection naturelle (sans formation de lignées) et de structurer de nouvelles recherches à leur égard. Un travail conceptuel, en amont de l'exploration empirique elle-même, serait donc nécessaire. En effet, il faut repenser les entités qui composent l'ontologie de la théorie de l'évolution par voie de sélection naturelle de sorte à les rendre elles aussi indépendantes de la notion de lignée. C'est seulement après un tel exercice théorique qu'il sera possible de coordonner la nouvelle ontologie évolutive avec des phénomènes et ainsi d'évaluer la possibilité de structurer la biologie de l'évolution autour d'une autre discipline que la phylogénétique.

Du point de vue de la philosophie de la biologie, des analyses épistémologiques par rapport à la biologie de l'évolution au-delà de la phylogénétique manquent toujours à l'appel. En effet, j'ai centré mon analyse sur la comparaison de deux itérations épistémiques de la

phylogénétique. Ce faisant, j'ai délaissé d'autres approches de la biologie de l'évolution et j'ai notamment contourné l'étude de la pensée des réseaux (*network thinking*). Cette approche, évoquée en 7.3.4, va beaucoup plus loin dans son importation des outils de la théorie des graphes que la phylogénétique centrée sur les réseaux. Il s'agit de mettre les explications topologiques à l'avant-plan de la biologie de l'évolution et de bien souligner que les outils de la théorie des graphes permettent d'appréhender les phénomènes biologiques sous différents angles. En effet, si les réseaux peuvent être utilisés pour modéliser les relations phylogénétiques, ils peuvent aussi servir à explorer d'autres types de relation entre les entités biologiques : liens de similarité (sans inférences phylogénétiques), relations fonctionnelles, interactions écologiques, cooccurrence, etc. Autrement dit, la pensée des réseaux modélise le monde vivant sur la base d'interactions entre les composantes d'un système et utilise la théorie des graphes pour explorer les caractéristiques de ces systèmes. Cela permet de mettre à jour des phénomènes qui pourraient autrement rester inaperçus, de souligner des interactions et des structures interactionnelles dont l'importance pourrait être capitale à la bonne compréhension des dynamiques biologiques (Bapteste et al. 2013; Corel et al. 2016; Corel et al. 2018; Watson et al. 2019; Watson, Habib et Bapteste 2020; Watson, Lopez et Bapteste 2021).

Cette manière de pratiquer la biologie de l'évolution diverge probablement grandement des pratiques phylogénétiques étudiées dans cette thèse. Nonobstant cette hypothèse, il conviendra, dans le futur, de réaliser une analyse épistémologique de cet ensemble de pratiques pour le resituer avec précision dans le reste de l'écosystème intellectuel qu'est la biologie de l'évolution. Outre le fait que ce chantier est immense et pourrait faire l'objet d'une thèse en soi, j'ai préféré ne pas me lancer dans une telle analyse parce que je considère certains de mes travaux comme étant trop proches de ce courant de pensée (Bapteste et Papale 2021; Papale, Saget et Bapteste 2020). Autrement dit, mes biais favorables à l'égard de la pensée des réseaux font de moi un candidat peu adéquat pour analyser épistémologiquement cette approche.

Dans une perspective de philosophie des sciences, le cohérentisme historique pourrait et devrait être confronté à d'autres phénomènes. Premièrement, il serait important, pour renforcer les constats de ma thèse, de démontrer l'efficacité du cohérentisme historique lorsqu'il est mobilisé pour analyser épistémologiquement d'autres pratiques classificatoires. En effet, mon

échantillon est plutôt limité (j'ai seulement considéré les pratiques phylogénétiques). J'ai toutefois confiance que des analyses similaires pourraient être réalisées pour les classifications biochimiques, chimiques, physiques, anthropologiques, médicales, etc. En effet, l'ensemble de ces pratiques partagent l'injonction à regrouper des entités au sein de catégories et de mettre en relation ces catégories. Or, les catégories sont inobservables directement, elles doivent être inférées sur la base des points communs entre les membres de ces catégories. La même chose peut être dite des liens entre catégories qui doivent être inférés sur la base d'éléments plus directement observables. Ainsi, je suggère que toute pratique classificatoire est traversée par le problème de la coordination et gagne donc à être étudiée dans une perspective cohérentiste et historique. Plus d'études pourront confirmer ou infirmer ce constat.

Deuxièmement, il serait extrêmement intéressant de tester la portée du cohérentisme historique au-delà des pratiques métrologiques et des pratiques classificatoires. Les descriptions qualitatives sont omniprésentes en sciences. Elles permettent de décrire les phénomènes avec précision, de donner un sens à des mesures ou à des classifications et de suggérer des pistes d'exploration pour d'éventuelles quantifications, classifications, etc. En plus d'être implicites à la pratique des sciences non humaines comme la physique, la chimie ou la biologie, elles sont explicitement au cœur d'autres disciplines comme l'anthropologie, la philosophie ou la psychologie. Nonobstant cette omniprésence, les descriptions qualitatives sont absentes des réflexions en philosophie des sciences. Cet angle mort gagne à être dissipé. Le cohérentisme historique pourrait bien offrir un cadre théorique pour amorcer la réflexion. En effet, le va-et-vient entre théorie et observations décrites de façon qualitative témoigne probablement du genre d'entrelacement intime qui requiert une méthode analytique rigoureuse pour être décomposé adéquatement.

Dans une perspective encore plus large, je propose que le cohérentisme historique puisse servir de cadre théorique de base pour l'unification des sciences sociales. J'ai souligné à quelques reprises que mon approche philosophique se concentrait de fait sur les aspects discursifs des pratiques scientifiques, mais que le cohérentisme historique invite à observer les pratiques scientifiques sous un maximum d'angles. L'observation participante typique de l'anthropologie, l'étude quantitative des relations au sein de la communauté (typique des sociologues et de la

bibliométrie), l'exploration historique des débats, l'étude de l'inscription des pratiques scientifiques dans une écologie intellectuelle plus englobante, voilà autant d'aspects que le cohérentisme historique nous exhorte à étudier. Une telle approche invite la pluridisciplinarité tout en offrant un modèle général au sein duquel les contributions respectives de chaque discipline pourraient contribuer à l'explication d'un seul et même phénomène : la transformation des systèmes épistémiques. Il ne s'agit pas de dire que les disciplines mentionnées pourraient être réduites à cette approche; il s'agit plutôt de souligner que le cohérentisme historique contribue à brouiller les frontières entre les pratiques scientifiques qui s'intéressent à différents aspects des mêmes phénomènes. Une telle proposition nécessite cependant une exploration philosophique et empirique plus approfondie.

Finalement, les avancées relatives à la compréhension des pratiques classificatoires réalisées dans cette thèse offrent des bases solides pour appréhender la dimension politique des gestes classificatoires. J'ai exploré dans cette thèse comment les descriptions classificatoires s'inscrivent au sein de discours aspirant au statut de discours gnoséologiques. Si ceux que j'ai étudiés sont des discours scientifiques, il va sans dire que des discours gnoséologiques existent bien au-delà des frontières de l'université : nos sociétés sont traversées par des régimes de connaissances, des aspirations à la vérité, qui viennent justifier nos actions individuelles et collectives. La présence de catégories au sein de ces discours est indéniable; des catégories comme celles associées au genre, au statut civique (citoyenne, résident permanent, résidente temporaire, non résident, etc.), à l'état matrimonial, à l'appartenance ethnique, à la criminalité, à l'âge, à la diplomation, à l'orientation sexuelle, parmi d'autres, structurent le vivre ensemble au sein de nos sociétés, que ces catégories soient institutionnalisées ou non. Comprendre la construction des discours gnoséologiques qui engendrent et mobilisent ces catégories est un projet souhaitable qui demande une philosophie réflexive, apte à décortiquer des schèmes de pensée bien ancrés. À travers le prisme du cohérentisme historique, la philosophie des sciences offre des outils qui favorisent l'analyse critique de nos croyances et des contraintes qu'elles imposent aux pratiques, que ces pratiques soient scientifiques ou sociales.

Références bibliographiques

- Acman, Mislav, Ruobing Wang, Lucy van Dorp, Liam P. Shaw, Qi Wang, Nina Luhmann, Yin Yuyao, Shijun Sun, Hongbin Chen et Hui Wang. 2021. « Role of the mobilome in the global dissemination of the carbapenem resistance gene *bla_{NDM}* ». *bioRxiv* 19 (15). <https://doi.org/10.1101/2021.01.14.426698>.
- Adanson, Michel. 1757. *Histoire naturelle du Sénégal : coquillages, avec la relation abrégée d'un voyage fait en ce pays, pendant les années 1749-53*. Paris : Bauche.
- . 1763. *Familles des plantes*. Paris : Vincent. <https://doi.org/10.5962/bhl.title.271>.
- Agüero-Chapin, Guillermin, Gisselle Pérez-Machado, Aminaél Sánchez-Rodríguez, Miguel Machado Santos et Agostinho Antunes. 2016. « Alignment-Free Methods for the Detection and Specificity Prediction of Adenylation Domains ». Dans *Nonribosomal Peptide and Polyketide Biosynthesis*. Sous la direction de Bradley S. Evans, 1401 : 253-72. New York : Springer New York. https://doi.org/10.1007/978-1-4939-3375-4_16.
- Altschul, Stephen F., Warren Gish, Webb Miller, Eugene W. Myers et David J. Lipman. 1990. « Basic local alignment search tool ». *Journal of molecular biology* 215 (3) : 403-10. [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2).
- Aouad, Monique, Guillaume Borrel, Céline Brochier-Armanet et Simonetta Gribaldo. 2019. « Evolutionary Placement of Methanonatronarchaeia ». *Nature Microbiology* 4 (4) : 558-59. <https://doi.org/10.1038/s41564-019-0359-z>.
- Aouad, Monique, Najwa Taib, Anne Oudart, Michel Lecocq, Manolo Gouy et Céline Brochier-Armanet. 2018. « Extreme Halophilic Archaea Derive from Two Distinct Methanogen Class II Lineages ». *Molecular Phylogenetics and Evolution* 127 : 46-54. <https://doi.org/10.1016/j.ympev.2018.04.011>.
- Ariew, André et Richard C. Lewontin. 2004. « The Confusions of Fitness ». *The British Journal for the Philosophy of Science* 55 (2) : 347-63. <https://doi.org/10.1093/bjps/55.2.347>.
- Assis, Leandro C. S. 2013. « Are Homology and Synapomorphy the Same or Different?: Letter to the Editor ». *Cladistics* 29 (1) : 7-9. <https://doi.org/10.1111/j.1096-0031.2012.00431.x>.
- . 2015. « Homology Assessment in Parsimony and Model-based Analyses: Two Sides of the Same Coin ». *Cladistics* 31 (3) : 315-20. <https://doi.org/10.1111/cla.12085>.
- Assis, Leandro C. S. et Ingo Brigandt. 2009. « Homology: Homeostatic Property Cluster Kinds in Systematics and Evolution ». *Evolutionary Biology* 36 (2) : 248-55. <https://doi.org/10.1007/s11692-009-9054-y>.
- Attias, H. 1999. « Inferring parameters and structure of latent variable models by variational Bayes ». Dans *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*. Sous la direction de Kathryn B. Laskey et Henri Prade, 21-30. San Francisco : Morgan Kaufmann Publishers Inc.
- Bapteste, Éric. 2007. « Au-delà de l'Arbre du vivant : pour une phylogénie postmoderne ». Thèse de doctorat, Université Paris 1.
- Bapteste, Éric et Yan Boucher. 2008. « Lateral gene transfer challenges principles of microbial systematics ». *Trends in microbiology* 16 (5) : 200-7. <https://doi.org/10.1016/j.tim.2008.02.005>.

- Bapteste, Eric, Yan Boucher, Jessica Leigh et W. Ford Doolittle. 2004. « Phylogenetic reconstruction and lateral gene transfer ». *Trends in microbiology* 12 (9) : 406-11. <https://doi.org/10.1016/j.tim.2004.07.002>.
- Bapteste, Eric et John Dupré. 2013. « Towards a processual microbial ontology ». *Biology & philosophy* 28 (2) : 379-404. <https://doi.org/10.1007/s10539-012-9350-2>.
- Bapteste, Eric et Philippe Huneman. 2018. « Towards a Dynamic Interaction Network of Life to Unify and Expand the Evolutionary Theory ». *BMC Biology* 16 (56). <https://doi.org/10.1186/s12915-018-0531-6>.
- Bapteste, Eric, Leo van Iersel, Axel Janke, Scot Kelchner, Steven Kelk, James O. McInerney, David A. Morrison et al. 2013. « Networks: expanding evolutionary thinking ». *Trends in Genetics* 29 (8) : 439-41. <https://doi.org/10.1016/j.tig.2013.05.007>.
- Bapteste, Eric, Maureen A. O'Malley, Robert G. Beiko, Marc Ereshefsky, J. Peter Gogarten, Laura Franklin-Hall, François-Joseph Lapointe et al. 2009. « Prokaryotic evolution and the tree of life are two different things ». *Biology direct* 4 (article 34). <https://doi.org/10.1186/1745-6150-4-34>.
- Bapteste, Eric et François Papale. 2021. « Modeling the Evolution of Interconnected Processes: It Is the Song and the Singers: Tracking Units of Selection with Interaction Networks ». *BioEssays* 43 (1) : 2000077. <https://doi.org/10.1002/bies.202000077>.
- Baquero, Fernando, Jose Luis Martínez, Val Fernandez Lanza, Jerónimo Rodríguez-Beltrán, Juan Carlos Galán, Alvaro San Millán, Rafael Cantón et Teresa M. Coque. 2021. « Evolutionary Pathways and Trajectories in Antibiotic Resistance ». *Clinical Microbiology Reviews* 34 (4) : e00050-19. <https://doi.org/10.1128/CMR.00050-19>.
- Barabási, A. L. 2016. *Network Science*. Boston : Cambridge University Press.
- Bernard, Charles, Romain Lannes, Yanyan Li, Eric Bapteste et Philippe Lopez. 2020. « Rich Repertoire of Quorum Sensing Protein Coding Sequences in CPR and DPANN Associated with Interspecies and Interkingdom Communication ». *mSystems* 5 (5) : e00414-20. <https://doi.org/10.1128/mSystems.00414-20>.
- Bernard, Guillaume, Cheong Xin Chan, Yao-ban Chan, Xin-Yi Chua, Yingnan Cong, James M. Hogan, Stefan R. Maetschke et Mark A. Ragan. 2017. « Alignment-Free Inference of Hierarchical and Reticulate Phylogenomic Relationships ». *Briefings in Bioinformatics* 20 (2) : 426-35. <https://doi.org/10.1093/bib/bbx067>.
- Bininda-Emonds, Olaf R. P. 2005. « Supertree Construction in the Genomic Age ». *Methods in Enzymology* 395 : 745-57. [https://doi.org/10.1016/S0076-6879\(05\)95038-6](https://doi.org/10.1016/S0076-6879(05)95038-6).
- Bird, Alexander. 2009. « Essences and Natural Kinds ». Dans *The Routledge Companion to Metaphysics*. Sous la direction de Robin Le Poidevin, Peter Simons, Andrew McGonigal et Ross P. Cameron, 497-506. Abingdon : Routledge.
- . 2010. « Discovering the Essences of Natural Kinds ». Dans *The Semantics and Metaphysics of Natural Kinds*. Sous la direction de Helen Beebe et Nigel Sabbarton-Leary, 125-36. Abingdon : Routledge.
- . 2018. « The Metaphysics of Natural Kinds ». *Synthese* 195 : 1397-1426. <https://doi.org/10.1007/s11229-015-0833-y>.
- Blackwell, Antoinette Louisa Brown. 1875. *The Sexes Throughout Nature*. New York : GP Putnam.

- Blanchette, Mathieu, Takashi Kunisawa et David Sankoff. 1999. « Gene order breakpoint evidence in animal mitochondrial phylogeny ». *Journal of Molecular Evolution* 49 (2) : 193-203. <https://doi.org/10.1007/pl00006542>.
- Bock, Walter J. 1963. « Evolution and Phylogeny in Morphologically Uniform Groups ». *The American Naturalist* 97 (896) : 265-85. <https://doi.org/10.1086/282280>.
- . 1973. « Philosophical Foundations of Classical Evolutionary Classification ». *Systematic Zoology* 22 (4) : 375. <https://doi.org/10.2307/2412945>.
- . 1977. « Foundations and methods of evolutionary classification ». Dans *Major Patterns in Vertebrate Evolution*. Sous la direction de Max K. Hecht, Peter C. Goody et Bessie M. Hecht, 851-95. New York : Plenum Press.
- Borrel, Guillaume, Panagiotis S. Adam, Luke J. McKay, Lin-Xing Chen, Isabel Natalia Sierra-García, Christian M. K. Sieber, Quentin Letourneur et al. 2019. « Wide Diversity of Methane and Short-Chain Alkane Metabolisms in Uncultured Archaea ». *Nature Microbiology* 4 (4) : 603-13. <https://doi.org/10.1038/s41564-019-0363-3>.
- Bouchard, Frédéric. 2008. « Causal processes, fitness, and the differential persistence of lineages ». *Philosophy of Science* 75 (5) : 560-70. <https://doi.org/10.1086/594507>.
- Bourrat, Pierrick. 2014. « From Survivors to Replicators: Evolution by Natural Selection Revisited ». *Biology & Philosophy* 29 (4) : 517-38. <https://doi.org/10.1007/s10539-013-9383-1>.
- Boyd, Robert N. 1989. « What realism implies and what it does not ». *Dialectica* 43 (1-2) : 5-29. <https://doi.org/10.1111/j.1746-8361.1989.tb00928.x>.
- . 1991. « Realism, anti-foundationalism and the enthusiasm for natural kinds ». *Philosophical studies* 61 (1) : 127-48. <https://doi.org/10.1007/BF00385837>.
- . 1997. « Kinds as the “workmanship of men”: Realism, constructivism, and natural kinds ». Dans *Rationality, realism, revision: Proceedings of the 3rd international congress of the Society for Analytical Philosophy*. Sous la direction de Julian Nida-Rumelin, 52-89. New York : Walter de Gruyter.
- . 1999a. « Homeostasis, species, and higher taxa ». Dans *Species: New Interdisciplinary Essays*. Sous la direction de R. N. Wilson, 141-85. Cambridge, MA : MIT Press.
- . 1999b. « Kinds, Complexity and Multiple realization ». *Philosophical Studies* 95 (1) : 67-98. <http://dx.doi.org/10.1023/A:1004511407133>.
- . 2013. « Semantic Externalism and Knowing Our Own Minds: Ignoring Twin-Earth and Doing Naturalistic Philosophy: Semantic Externalism and Knowing Our Own Minds ». *Theoria* 79 (3) : 204-28. <https://doi.org/10.1111/theo.12019>.
- Brigandt, Ingo. 2002. « Homology and the Origin of Correspondence ». *Biology & Philosophy* 17 (3) : 389-407. <https://doi.org/10.1023/A:1020196124917>.
- . 2003. « Homology in Comparative, Molecular, and Evolutionary Developmental Biology: The Radiation of a Concept ». *Journal of Experimental Zoology* 299B (1) : 9-17. <https://doi.org/10.1002/jez.b.36>.
- . 2007. « Typology Now: Homology and Developmental Constraints Explain Evolvability ». *Biology & Philosophy* 22 (5) : 709-25. <https://doi.org/10.1007/s10539-007-9089-3>.
- . 2009. « Natural Kinds in Evolution and Systematics: Metaphysical and Epistemological Considerations ». *Acta Biotheoretica* 57 (1-2) : 77-97. <https://doi.org/10.1007/s10441-008-9056-7>.

- . 2020. « How to Philosophically Tackle Kinds without Talking About 'Natural Kinds' ». *Canadian Journal of Philosophy*, 1-24. <https://doi.org/10.1017/can.2020.29>.
- Brigandt, Ingo et Paul E. Griffiths. 2007. « The Importance of Homology for Biology and Philosophy ». *Biology & Philosophy* 22 (5) : 633-41. <https://doi.org/10.1007/s10539-007-9094-6>.
- Brinkmann, Henner et Hervé Philippe. 1999. « Archaea Sister Group of Bacteria? Indications from Tree Reconstruction Artifacts in Ancient Phylogenies ». *Molecular Biology and Evolution* 16 (6) : 817-25. <https://doi.org/10.1093/oxfordjournals.molbev.a026166>.
- Brower, Andrew V. Z. 2000. « Evolution Is Not a Necessary Assumption of Cladistics ». *Cladistics* 16 (1) : 143-54. <https://doi.org/10.1111/j.1096-0031.2000.tb00351.x>.
- . 2015. « Transformational and Taxic Homology Revisited ». *Cladistics* 31 (2) : 197-201. <https://doi.org/10.1111/cla.12076>.
- Brower, Andrew V. Z. et Mario C. C. de Pinna. 2012. « Homology and Errors ». *Cladistics* 28 (5) : 529-38. <https://doi.org/10.1111/j.1096-0031.2012.00398.x>.
- Bryson, Vernon et Henry J. Vogel. 1965. *Evolving genes and proteins*. New York : Academic Press.
- Butler, Octavia E. 1993. *Parabol of the Sower*. New York : Four Walls Eight Windows.
- Byerly, Henry C. et Vincent A. Lazara. 1973. « Realist Foundations of Measurement ». *Philosophy of Science* 40 (1) : 10-28. <https://doi.org/10.1086/288493>.
- Campbell, Tayte P., Xiaoqing Sun, Vishal H. Patel, Crickette Sanz, David Morgan et Gautam Dantas. 2020. « The Microbiome and Resistome of Chimpanzees, Gorillas, and Humans across Host Lifestyle and Geography ». *The ISME Journal* 14 (6) : 1584-99. <https://doi.org/10.1038/s41396-020-0634-2>.
- Carine, Mark A. et Robert W. Scotland. 1999. « Taxic and Transformational Homology: Different Ways of Seeing ». *Cladistics* 15 (2) : 121-29. <https://doi.org/10.1111/j.1096-0031.1999.tb00254.x>.
- Carnap, Rudolph. 1950. *Logical foundations of probability*. Chicago : University of Chicago Press.
- . 1966. *Philosophical foundations of physics*. Réédité sous le titre *An Introduction to the Philosophy of Science*. Sous la direction de Martin Gardner. New York : Dover Publications.
- Cartwright, Nancy. 1999. *The dappled world: a study of the boundaries of science*. Cambridge : Cambridge University Press.
- Chang, Hasok. 1995. « Circularity and Reliability in Measurement ». *Perspectives on Science* 3 : 153-72.
- . 2001. « Spirit, air, and quicksilver: The search for the "real" scale of temperature ». *Historical studies in the physical and biological sciences* 31 (2) : 249-84. <http://dx.doi.org/10.1525/hsp.2001.31.2.249>.
- . 2004. *Inventing temperature: measurement and scientific progress*. Oxford studies in philosophy of science. Oxford : Oxford University Press.
- . 2007. « Scientific Progress: Beyond Foundationalism and Coherentism ». *Royal Institute of Philosophy Supplements* 61 : 1-20. <https://doi.org/10.1017/S1358246100009723>.
- . 2012. *Is water H₂O? evidence, realism and pluralism*. Boston studies in the philosophy of science, 293. New York : Springer Verlag.

- . 2016. « The rising of chemical natural kinds through epistemic iteration ». Dans *Natural kinds and classification in scientific practice*. Sous la direction de Catherine Kendig, 53-66. Londres : Routledge.
- . 2017. « Epistemic iteration and natural kinds: Realism and pluralism in taxonomy ». Dans *Philosophical Issues in Psychiatry IV: Psychiatric Nosology*. Sous la direction de Kenneth S. Kendler et Josef Parnas, 29. Oxford : Oxford University Press.
- Charbonneau, Mathieu. 2014. « Populations Without Reproduction ». *Philosophy of Science* 81 (5) : 727-40. <https://doi.org/10.1086/677203>.
- Charlebois, Robert L., Robert G. Beiko et Mark A. Ragan. 2003. « Microbial Phylogenomics: Branching Out ». *Nature* 421 (6920) : 217-217. <https://doi.org/10.1038/421217a>.
- Cherryh, Carolyne Janice. 1996. *Inheritor*. New York : DAW Books.
- . 2010. *Deceiver*. New York : DAW Books.
- Chiu, Lynn et Scott F. Gilbert. 2015. « The Birth of the Holobiont: Multi-Species Birthing Through Mutual Scaffolding and Niche Construction ». *Biosemiotics* 8 (2) : 191-210. <https://doi.org/10.1007/s12304-015-9232-5>.
- Churchland, Paul M. 1981. « Eliminative Materialism and the Propositional Attitudes ». *Journal of Philosophy* 78 (2) : 67-90. <https://doi.org/10.2307/2025900>.
- . 1988. *Matter and Consciousness*. Éd. rev. Cambridge, MA : MIT Press.
- Clavelin, Maurice. 2004. « Le copernicanisme et la mutation de la philosophie naturelle ». *Revue de métaphysique et de morale* 3 (43) : 353-70. <https://doi.org/10.3917/rmm.043.0353>.
- Cleland, Carol E. 2002. « Methodological and Epistemic Differences between Historical Science and Experimental Science* ». *Philosophy of Science* 69 (3) : 447-51. <https://doi.org/10.1086/342455>.
- Conix, Stijn et Pei-Shan Chi. 2021. « Against Natural Kind Eliminativism ». *Synthese* 198 : 8999-9020. <https://doi.org/10.1007/s11229-020-02614-2>.
- Conn, Christopher Hughes. 2002. « Locke on Natural Kinds and Essential Properties ». *Journal of Philosophical Research* 27 : 475-97. https://doi.org/10.5840/jpr_2002_12.
- Corel, Eduardo, Raphaël Meheust, Andrew K. Watson, James O. McInerney, Philippe Lopez et Eric Bapteste. 2018. « Bipartite Network Analysis of Gene Sharings in the Microbial World ». *Mol Biol Evol* 35 (4) : 899-913. <https://doi.org/10.1093/molbev/msy001>.
- Corel, Eduardo, Philippe Lopez, Raphaël Méheust et Eric Bapteste. 2016. « Network-Thinking: Graphs to Analyze Microbial Complexity and Evolution ». *Trends in Microbiology* 24 (3) : 224-37. <https://doi.org/10.1016/j.tim.2015.12.003>.
- Crane, Judith. 2003. « Locke's Theory of Classification ». *British Journal for the History of Philosophy* 11 (2) : 249-59. <https://doi.org/10.1080/0960878032000104859>.
- Craver, Carl F. 2009. « Mechanisms and natural kinds ». *Philosophical Psychology* 22 (5) : 575-94. <https://doi.org/10.1080/09515080903238930>.
- Crombie, Alistair Cameron. 1994. *Styles of scientific thinking in the European tradition: The history of argument and explanation especially in the mathematical and biomedical sciences and arts*. 3 vols. Londres : Duckworth.
- . 1995. « Commitments and styles of European scientific thinking ». *history of Science* 33 (2) : 225-38. <https://doi.org/10.1177%2F007327539503300204>.

- Dagan, Tal et William Martin. 2009. « Getting a Better Picture of Microbial Evolution En Route to a Network of Genomes ». *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1527) : 2187-96. <https://doi.org/10.1098/rstb.2009.0040>.
- Darwin, Charles A. 1859. *On the Origin of Species by Means of Natural Selection*. Londres : John Murray.
- Dawkins, Richard. 2006. *The Selfish Gene: 30th Anniversary edition*. Oxford : Oxford University Press.
- de Beer, Gavin. 1971. *Homology, an unsolved problem*. Vol. 11. Oxford : Oxford University Press.
- de Queiroz, Alan et John Gatesy. 2007. « The Supermatrix Approach to Systematics ». *Trends in Ecology & Evolution* 22 (1) : 34-41. <https://doi.org/10.1016/j.tree.2006.10.002>.
- de Queiroz, Kevin. 1988. « Systematics and the Darwinian revolution ». *Philosophy of Science* 55 (2) : 238-59. <https://doi.org/10.1086/289430>.
- . 1992. « Review of Principles of Systematic Zoology ». *Systematic Biology* 41 (2) : 264-66. <https://doi.org/10.2307/2992528>.
- . 1997. « The Linnaean hierarchy and the evolutionization of taxonomy, with emphasis on the problem of nomenclature ». *Aliso: A Journal of Systematic and Evolutionary Botany* 15 (2) : 125-44.
- . 2007. « Species Concepts and Species Delimitation ». *Systematic Biology* 56 (6) : 879-86. <https://doi.org/10.1080/10635150701701083>.
- Deleuze, Gilles. 1968. *Différence et répétition*. Paris : PUF.
- Delsuc, Frédéric, Henner Brinkmann et Hervé Philippe. 2005. « Phylogenomics and the Reconstruction of the Tree of Life ». *Nature Reviews Genetics* 6 (5) : 361-75. <https://doi.org/10.1038/nrg1603>.
- DeSalle, Riley. 2009. « The Twin Phylogenomic Challenges ». Dans *Darwin Evolution and Life*, 23-30. Incheon : NIBR Symposium.
- DeSalle, Rob et Margaret Riley. 2020. « Should Networks Supplant Tree Building? ». *Microorganisms* 8 (8) : 1179. <https://doi.org/10.3390/microorganisms8081179>.
- Devitt, Michael. 2007. « Referential Descriptions and Conversational Implicatures ». *European Journal of Analytic Philosophy* 3 (2) : 7-32.
- . 2008. « Resurrecting biological essentialism ». *Philosophy of Science* 75 (3) : 344-82. <https://doi.org/10.1086/593566>.
- . 2011. « Experimental Semantics ». *Philosophy and Phenomenological Research* 82 (2) : 418-35. <https://doi.org/10.1111/j.1933-1592.2010.00413.x>.
- Diez, José A. 1997. « A Hundred Years of Numbers. An Historical Introduction to Measurement Theory 1887-1990-Part II: Suppes and the Mature Theory. Representation and Uniqueness ». *Studies In History and Philosophy of Science Part A* 28 (2) : 237-65. [https://doi.org/10.1016/S0039-3681\(96\)00015-5](https://doi.org/10.1016/S0039-3681(96)00015-5).
- DiSalle, Robert. 2020. « Space and Time: Inertial Frames ». Dans *The Stanford Encyclopedia of Philosophy*. Éd. hiver 2020. Sous la direction d'Edward N. Zalta. Université de Stanford. <https://plato.stanford.edu/archives/win2020/entries/spacetime-iframe/>.
- Dobzhansky, Theodosius. 1973. « Nothing in Biology Makes Sense Except in the Light of Evolution ». *The American Biology Teacher* 35 (3) : 25-29. <https://doi.org/10.2307/4444260>.

- Doolittle, W. Ford. 1999. « Lateral Genomics ». *Trends in Biochemical Sciences* 24 (12) : M5-8.
[https://doi.org/10.1016/S0968-0004\(99\)01471-1](https://doi.org/10.1016/S0968-0004(99)01471-1).
- . 2000. « Uprooting the Tree of Life ». *Scientific American* 282 (2) : 90-95.
<https://doi.org/10.1038/scientificamerican0200-90>.
- . « The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis means for both of them ». *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 364 (1527) : 2221-28.
<https://doi.org/10.1098/rstb.2009.0032>.
- Doolittle, W. Ford et Éric Baptiste. 2007. « Pattern pluralism and the Tree of Life hypothesis ». *Proceedings of the National Academy of Sciences* 104 (7) : 2043-49.
<https://doi.org/10.1073/pnas.0610699104>.
- Doolittle, W. Ford et Austin Booth. 2017. « It's the Song, Not the Singer: An Exploration of Holobiosis and Evolutionary Theory ». *Biology & Philosophy* 32 (1) : 5-24.
<https://doi.org/10.1007/s10539-016-9542-2>.
- Doolittle, W. Ford et S. Andrew Inkpen. 2018. « Processes and Patterns of Interaction as Units of Selection: An Introduction to ITSNTS Thinking ». *Proceedings of the National Academy of Sciences of the United States of America* 115 (16) : 4006-14.
<https://doi.org/10.1073/pnas.1722232115>.
- Dupré, John. 1981. « Natural kinds and biological taxa ». *The Philosophical Review* 90 (1) : 66-90.
<https://doi.org/10.2307/2184373>.
- . 1993. *The disorder of things: Metaphysical foundations of the disunity of science*. Cambridge, MA : Harvard University Press.
- . 2012. *Processes of life: Essays in the philosophy of biology*. Oxford : Oxford University Press.
- Edgar, Robert C. 2004. « MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput ». *Nucleic Acids Research* 32 (5) : 1792-97.
<https://doi.org/10.1093/nar/gkh340>.
- Eisen, Jonathan A. et Claire M. Fraser. 2003. « Phylogenomics: Intersection of Evolution and Genomics ». *Science* 300 (5626) : 1706-7. <https://doi.org/10.1126/science.1086292>.
- Eldredge, Niles. 1979a. « Alternative approaches to evolutionary theory ». *Bulletin of Carnegie Museum of Natural History* 13: 7-19.
- . 1979b. « Cladism and common sense ». Dans *Phylogenetic analysis and paleontology*. Sous la direction de Joel Cracraft et Niles Eldredge, 165-98. New York : Columbia University Press.
- Elliott, Kevin C. 2011. « Direct and Indirect Roles for Values in Science* ». *Philosophy of Science* 78 (2) : 303-24. <https://doi.org/10.1086/659222>.
- . 2017. *A tapestry of values: an introduction to values in science*. Oxford : Oxford University Press.
- Ellis, Brian. 1966. *Basic Concepts of Measurement*. Cambridge : Cambridge University Press.
- . 2001. *Scientific essentialism*. Cambridge : Cambridge University Press.
- Ereshefsky, Marc. 1992. « Eliminative pluralism ». *Philosophy of Science* 59 (4), 671-90.
<https://doi.org/10.1086/289701>.
- . 2018. « Natural Kinds, Mind Independence, and Defeasibility ». *Philosophy of Science* 85 (5) : 845-56. <https://doi.org/10.1086/699676>.

- Ereshefsky, Marc et Makmiller Pedroso. 2015. « Rethinking evolutionary individuality ». *Proceedings of the National Academy of Sciences* 112 (33) : 10126-32. <https://doi.org/10.1073/pnas.1421377112>.
- Ereshefsky, Marc et Thomas A. C. Reydon. 2015. « Scientific kinds ». *Philosophical Studies* 172 (4) : 969-86. <https://doi.org/10.1007/s11098-014-0301-4>.
- Farris, James S. 1970. « Methods for Computing Wagner Trees ». *Systematic Biology* 19 (1) : 83-92. <https://doi.org/10.1093/sysbio/19.1.83>.
- . 1983. « The Logical basis of phylogenetic analysis ». Dans *Advances in Cladistics*. Sous la direction de Norman I. Platnick et Victoria Ann Funk, 2 : 7-36. New York : Columbia University Press.
- . 1990. « Haeckel, History, and Hull ». *Systematic Zoology* 39 (1) : 81. <https://doi.org/10.2307/2992211>.
- Felsenstein, J. 2004. *Inferring Phylogenies*. Sunderland, MA : Sinauer Associates.
- Feng, Yutian, Uri Neri, Sean Gosselin, Artemis S. Louyakis, R. Thane Papke, Uri Gophna et Johann Peter Gogarten. 2021. « The Evolutionary Origins of Extreme Halophilic Archaeal Lineages ». Sous la direction de Martin Embley. *Genome Biology and Evolution* 13 (8) : evab166. <https://doi.org/10.1093/gbe/evab166>.
- Fernández Moreno, Luis. 2016. *The Reference of Natural Kind Terms*. Berlin : Peter Lang Verlag.
- Ferreiro, Aura, Nathan Crook, Andrew J. Gasparrini et Gautam Dantas. 2018. « Multiscale evolutionary dynamics of host-associated microbiomes ». *Cell* 172 (6) : 1216-27. <https://dx.doi.org/10.1016%2Fj.cell.2018.02.015>.
- Fitch, Walter M. 2000. « Homology: A Personal View on Some of the Problems ». *Trends in Genetics* 16 (5) : 227-31. [https://doi.org/10.1016/S0168-9525\(00\)02005-9](https://doi.org/10.1016/S0168-9525(00)02005-9).
- Flores Espinola, Artemisa. 2013. « Science et Politique: quand le féminisme fait avancer la science ». *Raison présente* 186 (1) : 97-106. <https://doi.org/10.3406/raipr.2013.4469>.
- Fontaine, Naomi. 2019. *Shuni*. Montréal : Mémoire d'encrier.
- Forterre, Patrick. 2012. « Darwin's goldmine is still open: variation and selection run the world ». *Frontiers in Cellular and Infection Microbiology* 2 : 106. <https://doi.org/10.3389/fcimb.2012.00106>.
- Foucault, Michel. 1966. *Les Mots et les choses*. Paris : Éditions Gallimard.
- Francis, Andrew R. et Mike Steel. 2015. « Which Phylogenetic Networks Are Merely Trees with Additional Arcs? ». *Systematic Biology* 64 (5) : 768-77. <https://doi.org/10.1093/sysbio/syv037>.
- Franklin-Hall, Laura R. 2015. « Natural Kinds as Categorical Bottlenecks ». *Philosophical Studies* 172 (4) : 925-48. <https://doi.org/10.1007/s11098-014-0326-8>.
- Frigerio, Aldo, Alessandro Giordani et Luca Mari. 2010. « Outline of a General Model of Measurement ». *Synthese* 175 (2) : 123-49. <https://doi.org/10.1007/s11229-009-9466-3>.
- Gayon, Jean. 1992. *Darwin et l'après Darwin : une histoire de l'hypothèse de sélection naturelle. Histoire des idées, théorie politique et recherches en sciences sociales*. Paris : Éditions Kimé.
- Geisler, Jonathan H., Michael R. McGowen, Guang Yang et John Gatesy. 2011. « A supermatrix analysis of genomic, morphological, and paleontological data from crown Cetacea ». *BMC evolutionary biology* 11 (1) : 1-33. <https://doi.org/10.1186/1471-2148-11-112>.

- Ghaly, Timothy M. et Michael R. Gillings. 2018. « Mobile DNAs as Ecologically and Evolutionarily Independent Units of Life ». *Trends in Microbiology* 26 (11) : 904-12.
<https://doi.org/10.1016/j.tim.2018.05.008>.
- Ghiselin, Michael T. 1974. « A radical solution to the species problem ». *Systematic Biology* 23 (4) : 536-44. <https://doi.org/10.1093/sysbio/23.4.536>.
- Gilbert, Scott F. et Alfred I. Tauber. 2016. « Rethinking Individuality: The Dialectics of the Holobiont ». *Biology & Philosophy* 31 (6) : 839-53. <https://doi.org/10.1007/s10539-016-9541-3>.
- Godfrey-Smith, Peter. 2009. *Darwinian Populations and Natural Selection*. Oxford : Oxford University Press.
- Goldenfeld, Nigel et Carl Woese. 2007. « Biology's next Revolution ». *Nature* 445 (7126) : 369-369. <https://doi.org/10.1038/445369a>.
- Gontier, Nathalie. 2011. « Depicting the Tree of Life: The Philosophical and Historical Roots of Evolutionary Tree Diagrams ». *Evolution: Education and Outreach* 4 (3) : 515-38.
<https://doi.org/10.1007/s12052-011-0355-0>.
- Gould, Stephen Jay. 1985. « A clock of evolution ». *Journal of Natural History* 4 : 12-25.
- Graham, David W. Charles W. Knapp, Bent T. Christensen, Seánín McCluskey et Jan Dolfing. 2016. « Appearance of β -Lactam Resistance Genes in Agricultural Soils and Clinical Isolates over the 20th Century ». *Scientific Reports* 6 (1) : 21550.
<https://doi.org/10.1038/srep21550>.
- Grant, Taran et Arnold G. Kluge. 2004. « Transformation Series as an Ideographic Character Concept ». *Cladistics* 20 (1) : 23-31. <https://doi.org/10.1111/j.1096-0031.2004.00003.x>.
- Grant, Verne E. 1971. *Plant Speciation*. New York : Columbia University Press.
- Griesemer, James R. 2000. « The Units of Evolutionary Transition ». *Selection* 1 (1-3) : 67-80.
<https://doi.org/10.1556/Select.1.2000.1-3.7>.
- Griffiths, Paul E. 1999. « Squaring the Circle: Natural Kinds with Historical Essences ». Dans *Species: New Interdisciplinary Essays*. Sous la direction de Robert A. Wilson, 209-28. Cambridge, MA : MIT Press.
- . 2007. « The Phenomena of Homology ». *Biology & Philosophy* 22 (5) : 643-58.
<https://doi.org/10.1007/s10539-007-9090-x>.
- Guindon, Stéphane, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk et Olivier Gascuel. 2010. « New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0 ». *Systematic Biology* 59 (3) : 307-21. <https://doi.org/10.1093/sysbio/syq010>.
- Hacking, Ian. 1991. « A tradition of natural kinds ». *Philosophical Studies* 61 (1) : 109-26.
- . 1992. « 'Style' for historians and philosophers ». *Studies in History and Philosophy of Science Part A* 23 (1) : 1-20. [https://doi.org/10.1016/0039-3681\(92\)90024-Z](https://doi.org/10.1016/0039-3681(92)90024-Z).
- . 1994. « Styles of Scientific Thinking or Reasoning: A New Analytical Tool for Historians and Philosophers of the Sciences ». Dans *Trends in the Historiography of Science*. Sous la direction de Kostas Gavroglu, Jean Christianidis et Efthymios Nicolaidis, 31-48. Boston Studies in the Philosophy of Science. Dordrecht : Springer. https://doi.org/10.1007/978-94-017-3596-4_3.
- . 1999. *The social construction of what?* Cambridge, MA : Harvard University Press.

- . 2007a. « Natural kinds: Rosy dawn, scholastic twilight ». *Royal Institute of Philosophy Supplement* 61 : 203-39. <https://doi.org/10.1017/S1358246100009802>.
- . 2007b. « Putnam's theory of natural kinds and their names is not the same as Kripke's ». *Principia* 11 (1) : 1. <https://doi.org/10.5007/%25x>.
- . 2009. *Scientific reason*. Taipei : National Taiwan University Press.
- Haggerty, Leanne S., Pierre-Alain Jachiet, William P. Hanage, David A. Fitzpatrick, Philippe Lopez, Mary J. O'Connell, Davide Pisani, Mark Wilkinson, Eric Bapteste et James O. McInerney. 2014. « A Pluralistic Account of Homology: Adapting the Models to the Data ». *Molecular Biology and Evolution* 31 (3) : 501-16. <https://doi.org/10.1093/molbev/mst228>.
- Häggqvist, Sören et Åsa Wikforss. 2017. « Natural Kinds and Natural Kind Terms: Myth and Reality ». *The British Journal for the Philosophy of Science* 69 (4) : 911-33. <http://dx.doi.org/10.1093/bjps/axw041>.
- Halfon, Marc S. 2017. « Perspectives on Gene Regulatory Network Evolution ». *Trends in Genetics: TIG* 33 (7) : 436-47. <https://doi.org/10.1016/j.tig.2017.04.005>.
- Hall, B. 2007. « Homoplasmy and Homology: Dichotomy or Continuum? ». *Journal of Human Evolution* 52 (5) : 473-79. <https://doi.org/10.1016/j.jhevol.2006.11.010>.
- Haraoui, Louis-Patrick. 2021. « Networked Collective Microbiomes and the Rise of Subcellular "Units of Life" ». *Trends in Microbiology* S0966-842X2 (21) : 00238-9. <https://doi.org/10.1016/j.tim.2021.09.011>.
- Haslanger, Sally. 1995. « Ontology and Social Construction ». *Philosophical Topics* 23 (2) : 95-125. <https://doi.org/10.5840/philtopics19952324>.
- Havstad, Joyce C. 2017. « Messy Chemical Kinds ». *The British Journal for the Philosophy of Science* 69 (3) : 719-73. <https://doi.org/10.1093/bjps%2Faxw040>.
- Hazelwood, Caleb. 2017. « Synthetic Biology and Natural Kinds: Homeostatic Property Cluster Theory in the Post-Genomic Era. ». *Stance (1943-1880)* 10 : 49-57. <https://doi.org/10.5840/stance2017106>.
- Hennig, Willi. 1965. « Phylogenetic systematics ». *Annual review of entomology* 10 (1) : 97-116. <https://doi.org/10.1146/annurev.en.10.010165.000525>.
- . 1966. *Phylogenetic Systematics*. Traduit par D. Dwight Davis et Rainer Zangerl. Urbana : University of Illinois Press.
- . 1975. « "Cladistic Analysis or Cladistic Classification?": A Reply to Ernst Mayr ». *Systematic Zoology* 24 (2) : 244. <https://doi.org/10.2307/2412765>.
- . 1981. *Insect Phylogeny*. Traduit par Adrian Charles Pont. Chichester, R.-U : J. Wiley.
- Higgins, Desmond G. et Paul M. Sharp. 1988. « CLUSTAL: A Package for Performing Multiple Sequence Alignment on a Microcomputer ». *Gene* 73 (1) : 237-44. [https://doi.org/10.1016/0378-1119\(88\)90330-7](https://doi.org/10.1016/0378-1119(88)90330-7).
- Hoborn, Donald. 2021. « BIOSCAN: DNA Barcoding to Accelerate Taxonomy and Biogeography for Conservation and Sustainability ». *Genome* 64 (3) : 161-64. <https://doi.org/10.1139/gen-2020-0009>.
- Hoquet, Thierry. 2005. *Buffon: histoire naturelle et philosophie*. Paris : Champion.
- Hrdy, Sarah Blaffer. 1999. *The woman that never evolved: with a new preface and bibliographical updates*. Éd. rev. Cambridge, MA : Harvard University Press.

- Hu, Yongfei, George F. Gao et Baoli Zhu. 2017. « The Antibiotic Resistome: Gene Flow in Environments, Animals and Human Beings ». *Frontiers of Medicine* 11 (2) : 161-68. <https://doi.org/10.1007/s11684-017-0531-x>.
- Hu, Yongfei, Xi Yang, Jing Li, Na Lv, Fei Liu, Jun Wu, Ivan Y. C. Lin et al. 2016. « The Bacterial Mobile Resistome Transfer Network Connecting the Animal and Human Microbiomes ». *Applied and Environmental Microbiology* 82 (22) : 6672-81. <https://doi.org/10.1128/AEM.01802-16>.
- Hull, David L. 1965a. « The Effect of Essentialism on Taxonomy--Two Thousand Years of Stasis (I) ». *The British Journal for the Philosophy of Science* 15 (60) : 314-26. <https://doi.org/10.1093/bjps/XV.60.314>.
- . 1965b. « The Effect of Essentialism on Taxonomy--Two Thousand Years of Stasis (II) ». *The British Journal for the Philosophy of Science* 16 (61) : 1-18. <https://doi.org/10.1093/bjps/XVI.61.1>.
- . 1970. « Contemporary systematic philosophies ». *Annual Review of ecology and systematics* 1 (1) : 19-54. <https://doi.org/10.1146/annurev.es.01.110170.000315>.
- . 1976. « Are species really individuals? » *Systematic Biology* 25 (2) : 174-91. <https://doi.org/10.2307/2412744>.
- . 1988. *Science as a Process*. Chicago : Chicago University Press.
- . 2001. « The Role of Theories in Biological Systematics ». *Studies in History and Philosophy of Biol & Biomed Sci* 32 (2) : 221-38. [https://doi.org/10.1016/S1369-8486\(01\)00006-1](https://doi.org/10.1016/S1369-8486(01)00006-1).
- Huneman, Philippe. 2010. « Topological Explanations and Robustness in Biological Sciences ». *Synthese* 177 (2) : 213-45. <https://doi.org/10.1007/s11229-010-9842-z>.
- Hurley, Kameron. 2011. *Infidel*. San Francisco : Night Shade Books.
- Huson, Daniel H. 2007. « Split Networks and Reticulate Networks ». Dans *Reconstructing evolution: new mathematical and computational advances*. Sous la direction de Olivier Gascuel et Mike A. Steel, 247-76. Oxford : Oxford University Press.
- Huson, Daniel H., Tobias Klöpper, Pete J. Lockhart et Mike A. Steel. 2005. « Reconstruction of Reticulate Networks from Gene Trees ». Dans *Research in Computational Molecular Biology*. Lecture Notes in Computer Science, vol. 3500. Sous la direction de Satoru Miyano, Jill Mesirov, Simon Kasif, Sorin Istrail, Pavel A. Pevzner et Michael Waterman, 233-49. Berlin : Springer. https://doi.org/10.1007/11415770_18.
- Huson, Daniel H. et Celine Scornavacca. 2011. « A Survey of Combinatorial Methods for Phylogenetic Networks ». *Genome Biology and Evolution* 3 (janvier) : 23-35. <https://doi.org/10.1093/gbe/evq077>.
- Huson, Daniel, Regula Rupp et Celine Scornavacca. 2010. *Phylogenetic Networks: Concepts, Algorithms and Applications*. Cambridge : Cambridge University Press.
- Inkpen, S. Andrew et W. Ford Doolittle. 2016. « Molecular Phylogenetics and the Perennial Problem of Homology ». *Journal of Molecular Evolution* 83 (5-6) : 184-92. <https://doi.org/10.1007/s00239-016-9766-4>.
- Intemann, Kristen. 2015. « Distinguishing between legitimate and illegitimate values in climate modeling ». *European Journal for Philosophy of Science* 5 (2) : 217-32. <https://doi.org/10.1007/s13194-014-0105-6>.

- Iranzo, Jaime, Eugene V. Koonin, David Prangishvili et Mart Krupovic. 2016. « Bipartite Network Analysis of the Archaeal Virosphere: Evolutionary Connections between Viruses and Capsidless Mobile Elements ». *Journal of Virology* 90 (24) : 11043-55. <https://doi.org/10.1128/JVI.01622-16>.
- Iranzo, Jaime, Mart Krupovic et Eugene V. Koonin. 2016. « The Double-Stranded DNA Virosphere as a Modular Hierarchical Network of Gene Sharing ». *MBio* 7 (4) : e00978-16. <https://doi.org/10.1128/mBio.00978-16>.
- Isaac, Alistair M. C. 2019. « Epistemic Loops and Measurement Realism ». *Philosophy of Science* 86 (5) : 930-41. <https://doi.org/10.1086/705476>.
- Israel-Jost, Vincent. 2019. « Imagerie et sciences du vivant : une approche épistémologique ». Dans *Rencontres de Venise - regards sur l'épistémologie à l'intention des curieux en sciences du vivant*. Sous la direction de Zouaoui Boudjeltia et Michel Vanhaeberbeek, 157-83. Paris : Hermann Éditeurs.
- Jardine, Nicholas. 1969. « The Observational and Theoretical Components of Homology: A Study Based on the Morphology of the Dermal Skull-Roofs of Rhipidistian Fishes ». *Biological Journal of the Linnean Society* 1 (4) : 327-61. <https://doi.org/10.1111/j.1095-8312.1969.tb00125.x>.
- Jemisin, Nora Keita. 2015. *The Fifth Season*. Londres : Orbit.
- Jin, Guohua, Luay Nakhleh, Sagi Snir et Tamir Tuller. 2007. « Inferring Phylogenetic Networks by the Maximum Parsimony Criterion: A Case Study ». *Molecular Biology and Evolution* 24 (1) : 324-37. <https://doi.org/10.1093/molbev/msl163>.
- Jones, Jan-Erik. 2016. « Locke on Real Essence ». Dans *The Stanford Encyclopedia of Philosophy*. Sous la direction d'Edward N. Zalta. Université de Stanford.
- Kapli, Paschalia, Ziheng Yang et Maximilian J. Telford. 2020. « Phylogenetic Tree Building in the Genomic Age ». *Nature Reviews Genetics* 21 (7) : 428-44. <https://doi.org/10.1038/s41576-020-0233-0>.
- Kawashima, Takeshi, Shuichi Kawashima, Chisaki Tanaka, Miho Murai, Masahiko Yoneda, Nicholas H. Putnam, Daniel S. Rokhsar, Minoru Kanehisa, Nori Satoh et Hiroshi Wada. 2009. « Domain shuffling and the evolution of vertebrates ». *Genome Res* 19 (8) : 1393-1403. <https://doi.org/10.1101/gr.087072.108>.
- Kearney, Maureen et Olivier Rieppel. 2006. « Rejecting "the given" in Systematics ». *Cladistics* 22 (4) : 369-77. <https://doi.org/10.1111/j.1096-0031.2006.00110.x>.
- Keeling, Patrick J. et Jeffrey D. Palmer. 2008. « Horizontal gene transfer in eukaryotic evolution ». *Nat Rev Genet* 9 (8) : 605-18. <https://doi.org/10.1038/nrg2386>.
- Keller, Evelyn Fox. 2014. « From gene action to reactive genomes ». *The Journal of Physiology* 592 (11) : 2423-29. <https://dx.doi.org/10.1113%2Fjphysiol.2014.270991>.
- Keller, Roberto A., Richard N. Boyd et Quentin D. Wheeler. 2003. « The Illogical Basis of Phylogenetic Nomenclature ». *The Botanical Review* 69 (1) : 93-110. <https://www.jstor.org/stable/4354448>.
- Kendig, Catherine E., 2016a. « Homologizing as Kinding ». Dans *Natural Kinds and Classification in Scientific Practice*. Sous la direction de Catherine E. Kendig, 106-25. New York : Routledge.
- . 2016b. *Natural kinds and classification in scientific practice*. New York : Routledge.

- Khalidi, Muhammad Ali. 2013. *Natural categories and human kinds: Classification in the natural and social sciences*. Cambridge : Cambridge University Press.
- . 2018. « Natural Kinds as Nodes in Causal Networks ». *Synthese* 195 (4) : 1379-96. <https://doi.org/10.1007/s11229-015-0841-y>.
- Kluge, Arnold G. 1989. « A Concern for Evidence and a Phylogenetic Hypothesis of Relationships among Epicrates (Boidae, Serpentes) ». *Systematic Biology* 38 (1) : 7-25. <https://doi.org/10.1093/sysbio/38.1.7>.
- . 1997. « Testability and the Refutation and Corroboration of Cladistic Hypotheses ». *Cladistics* 13 (1-2) : 81-96. <https://doi.org/10.1111/j.1096-0031.1997.tb00242.x>.
- . 1998. « Total Evidence Or Taxonomic Congruence: Cladistics Or Consensus Classification ». *Cladistics* 14 (2) : 151-58. <https://doi.org/10.1006/clad.1997.0056>.
- . 2002. « Distinguishing “or” from “and” and the Case for Historical Identification ». *Cladistics* 18 (6) : 585-93. <https://doi.org/10.1111/j.1096-0031.2002.tb00294.x>.
- . 2003. « The Repugnant and the Mature in Phylogenetic Inference: Atemporal Similarity and Historical Identity ». *Cladistics* 19 (4) : 356-68. <https://doi.org/10.1111/j.1096-0031.2003.tb00379.x>.
- Kluge, Arnold G., et J. S. Farris. 1969. « Quantitative Phyletics and the Evolution of Anurans ». *Systematic Biology* 18 (1) : 1-32. <https://doi.org/10.1093/sysbio/18.1.1>.
- Knapp, Charles W., Jan Dolfing, Phillip Al Ehlert et David W. Graham. 2010. « Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940 ». *Environmental science & technology* 44 (2) : 580-87.
- Koonin, Eugene V. 2009a. « Towards a Postmodern Synthesis of Evolutionary Biology ». *Cell Cycle* 8 (6) : 799-800. <https://doi.org/10.4161/cc.8.6.8187>.
- . 2009b. « The Origin at 150: Is a New Evolutionary Synthesis in Sight? » *Trends in Genetics* 25 (11) : 473-75. <https://doi.org/10.1016/j.tig.2009.09.007>.
- Koonin, Eugene V., Kira S. Makarova et Yuri I. Wolf. 2021. « Evolution of Microbial Genomics: Conceptual Shifts over a Quarter Century ». *Trends in Microbiology* 29 (7) : 582-92. <https://doi.org/10.1016/j.tim.2021.01.005>.
- Koonin, Eugene V. et Yuri I. Wolf. 2009. « Is evolution Darwinian or/and Lamarckian? » *Biology direct* 4 (1) : 42. <https://doi.org/10.1186/1745-6150-4-42>.
- Kostka, Martin, Magdalena Uzlikova, Ivan Cepicka et Jaroslav Flegr. 2008. « SlowFaster, a User-Friendly Program for Slow-Fast Analysis and Its Application on Phylogeny of Blastocystis ». *BMC Bioinformatics* 9 (1) : 341. <https://doi.org/10.1186/1471-2105-9-341>.
- Kripke, Saul A. 1972. « Naming and necessity ». Dans *Semantics of natural language*, 253-355. Berlin : Springer.
- Lane, Christopher E. et John M. Archibald. 2008. « The eukaryotic tree of life: endosymbiosis takes its TOL ». *Trends Ecol Evol* 23 (5) : 268-75. <https://doi.org/10.1016/j.tree.2008.02.004>.
- LaPorte, J. 2004. *Natural Kinds and Conceptual Change*. Cambridge : Cambridge University Press.
- Latour, Bruno. 1987. *Science in Action*. Cambridge, MA : Harvard University Press.
- . 2003. « The Promises of Constructivism ». Dans *Chasing Technology: Matrix of Materiality*. Sous la direction de Don Ihde et Evan Selinger, 27-46. Indiana Series for the Philosophy of Science. Indiana : Indiana University Press.

- . 2015. *Face à Gaïa: huit conférences sur le nouveau régime climatique*. Paris : La Découverte.
- Le Guin, Ursula K. 1969. *The Left Hand of Darkness*. New York : Ace Books.
- Le Guyader, Hervé. 2001. « Qu'est-ce qu'un gène ? Une petite histoire du concept ». *Le courrier de l'environnement de l'INRA* (44) : 53-63.
- Lecointre, Guillaume et Philippe Huneman. 2020. « Que signifie « se ressembler » en biologie? » *Philosophia Scientiae* 24 (2) : 75-98. <https://doi.org/10.4000/philosophiascientiae.2304>.
- Lenton, Timothy J. 1998. « Gaia and natural selection ». *Nature* 394 (6692) : 439-47. <https://doi.org/10.1038/28792>.
- Lewontin, Richard C. 1970. « The Units of Selection ». *Annual Review of Ecology and Systematics* 1 : 1-18. <https://doi.org/10.1146/annurev.es.01.110170.000245>.
- Linné, Carl. 1751. *Philosophia botanica*. Amsterdam : Z. Chatelain.
- Lloyd, Elisabeth A. 1995. « Objectivity and the double standard for feminist epistemologies ». *Synthese* 104 (3) : 351-81. <https://doi.org/10.1007/BF01064505>.
- . 2018. « Holobionts as units of selection: holobionts as interactors, reproducers, and manifestors of adaptation ». Dans *Landscapes of collectivity in the life sciences*. Sous la direction de Snait B. Gissis, Ehud Lamm et Ayelet Shavit, 351-68. Cambridge, MA : MIT Press.
- Locke, John. 1975. *An Essay Concerning Human Understanding*. Sous la direction de Peter H. Nidditch. Oxford : Clarendon Press.
- Logan, Latania K. et Robert A. Weinstein. 2017. « The epidemiology of carbapenem-resistant Enterobacteriaceae: the impact and evolution of a global menace ». *The Journal of infectious diseases* 215 (suppl_1) : S28-36. <https://doi.org/10.1093/infdis/jiw282>.
- Longino, Helen E. 1990. *Science as social knowledge: values and objectivity in scientific inquiry*. Princeton : Princeton University Press.
- . 2016. « The Social Dimensions of Scientific Knowledge ». Dans *The Stanford Encyclopedia of Philosophy*. Éd. printemps 2016. Sous la direction d'Edward N. Zalta. Université de Stanford. <http://plato.stanford.edu/archives/spr2016/entries/scientific-knowledge-social/>.
- Losos, Jonathan B., David M. Hillis et Harry W. Greene. 2012. « Who Speaks with a Forked Tongue? » *Science* 338 (6113) : 1428-29. <https://doi.org/10.1126/science.1232455>.
- Love, Alan C. 2007. « Functional Homology and Homology of Function: Biological Concepts and Philosophical Consequences ». *Biology & Philosophy* 22 (5) : 691-708. <https://doi.org/10.1007/s10539-007-9093-7>.
- Lovejoy, Arthur O. 1936. *The great chain of being: a study of the history of an idea*. Cambridge, MA : Harvard University Press.
- Lowe, Edward Jonathan. 2011. « Locke on Real Essence and Water as a Natural Kind: A Qualified Defence ». *Aristotelian Society Supplementary Volume* 85 (1) : 1-19. <https://doi.org/10.1111/j.1467-8349.2011.00193.x>.
- Luce, R. Duncan, David H. Krantz, Patrick Suppes et Amos Tversky. 1990. *Foundations of Measurement*. Vol. 3. San Diego : Academic Press.
- Ludwig, David. 2018. « Letting Go of "Natural Kind": Toward a Multidimensional Framework of Nonarbitrary Classification ». *Philosophy of Science* 85 (1) : 31-52. <https://doi.org/10.1086/694835>.

- Mach, Ernst. 1896. « Critique of the Concept of Temperature ». Dans *Basic Concepts of Measurement*. Sous la direction de Brian Ellis, 183-96. Cambridge : Cambridge University Press.
- Macleay, William Sharp. 1821. *Horae Entomologicae: or Essays on the Annulose Animals, Vol. 1, Part 2. Containing an Attempt to Ascertain the Rank and Situation which the Celebrated Insect, Scarabeus Sacer, Holds Among Organised Beings*. Londres : S. Bagster.
- MacLeod, Miles et Thomas A. C. Reydon. 2013. « Natural Kinds in Philosophy and in the Life Sciences: Scholastic Twilight or New Dawn? » *Biological Theory* 7 (2) : 89-99. <https://doi.org/10.1007/s13752-012-0080-0>.
- Magnus, Paul D. 2014a. « Epistemic categories and causal kinds ». *Philosophy Faculty Scholarship*. 22.
- . 2014b. « NK≠HPC ». *The Philosophical Quarterly* 64 (256) : 471-77. <https://doi.org/10.1093/pq/pqu010>.
- . 2014c. « No grist for Mill on natural kinds ». *Journal for the History of Analytical Philosophy* 2 (4) : 1-15.
- . 2015a. « Taxonomy, ontology, and natural kinds ». *Synthese* 195 : 1427-1439. <https://doi.org/10.1007/s11229-015-0785-2>.
- . 2015b. « John Stuart Mill on Taxonomy and Natural Kinds ». *HOPOS: The Journal of the International Society for the History of Philosophy of Science* 5 (2) : 269-80. <https://doi.org/10.1086/682373>.
- Makarenkov, Vladimir et Pierre Legendre. 2000. « Improving the additive tree representation of a given dissimilarity matrix using reticulation ». Dans *Data analysis, classification, and related methods*. Sous la direction de Henk A. L. Kiers, Jean-Paul Rasson, Patrick J. F. Groenen et Martin Schade, 35-46. Berlin : Springer.
- Mallet, James. 2005. « Hybridization as an Invasion of the Genome ». *Trends in Ecology & Evolution* 20 (5) : 229-37. <https://doi.org/10.1016/j.tree.2005.02.010>.
- Mallet, James, Nora Besansky et Matthew W. Hahn. 2016. « How Reticulated Are Species? » *BioEssays* 38 (2) : 140-49. <https://doi.org/10.1002/bies.201500149>.
- Mari, Luca. 2000. « Beyond the Representational Viewpoint: A New Formalization of Measurement ». *Measurement* 27 (2) : 71-84. [https://doi.org/10.1016/S0263-2241\(99\)00055-X](https://doi.org/10.1016/S0263-2241(99)00055-X).
- . 2005. « The Problem of Foundations of Measurement ». *Measurement* 38 (4) : 259-66. <https://doi.org/10.1016/j.measurement.2005.09.006>.
- Mari, Luca et Alessandro Giordani. 2012. « Quantity and quantity value ». *Metrologia* 49 (6) : 756-64. <https://doi.org/10.1088/0026-1394/49/6/756>.
- Martijn, Joran, Max E. Schön, Anders E. Lind, Julian Vosseberg, Tom A. Williams, Anja Spang et Thijs J. G. Ettema. 2020. « Hikarchaeia Demonstrate an Intermediate Stage in the Methanogen-to-Halophile Transition ». *Nature Communications* 11 (1) : 5490. <https://doi.org/10.1038/s41467-020-19200-2>.
- Martin, William F., Sriram Garg et Verena Zimorski. 2015. « Endosymbiotic theories for eukaryote origin ». *Philos Trans R Soc Lond B Biol Sci* 370 (1678) : 20140330. <https://doi.org/10.1098/rstb.2014.0330>.
- Martinez, Eduardo J. 2017. « Stable Property Clusters and Their Grounds ». *Philosophy of Science* 84 (5) : 944-955. <https://doi.org/10.1086/693988>.

- Martínez, Manolo. 2020. « Synergic Kinds ». *Synthese* 197 (5) : 1931-46.
<https://doi.org/10.1007/s11229-017-1480-2>.
- Matthen, Mohan et André Ariew. 2009. « Selection and Causation* ». *Philosophy of Science* 76 (2) : 201-24.
- Mayr, Ernst. 1959. « Darwin and the evolutionary theory in biology ». Dans *Evolution and anthropology: A centennial appraisal*. Sous la direction de Betty J. Meggers, 1-10. Washington, D.C. : Anthropological Society of Washington.
- . 1969a. *Principles of Systematic Zoology*. New York: McGraw-Hill.
- . 1969b. « The Biological Meaning of Species ». *Biological Journal of the Linnean Society* 1 (3) : 311-20. <https://doi.org/10.1111/j.1095-8312.1969.tb00123.x>.
- . 1974. « Cladistic Analysis or Cladistic Classification? » *Zeitschrift Für Zoologische Systematik Und Evolutionsforschung* 12 (1) : 94-128. <https://doi.org/10.1111/j.1439-0469.1974.tb00160.x>.
- . 1981. « Biological Classification: Toward a Synthesis of Opposing Methodologies ». *Science* 214 (4520) : 510-16. <https://doi.org/10.1126/science.214.4520.510>.
- . 1994. « Typological versus population thinking ». Dans *Conceptual issues in evolutionary biology*. Sous la direction de Elliott Sober, 157-60. Cambridge, MA : MIT Press.
- McClintock, Barbara. 1984. « The Significance of Responses of the Genome to Challenge ». *Science* 226 (4676) : 792-801. <https://doi.org/10.1126/science.15739260>.
- McDaniel, Lauren D., Elizabeth Young, Jennifer Delaney, Fabian Ruhnau, Kim B. Ritchie et John H. Paul. 2010. « High frequency of horizontal gene transfer in the oceans ». *Science* 330 (6000) : 50. <https://doi.org/10.1126/science.1192243>.
- Méheust, Raphaël, Philippe Lopez et Eric Bapteste. 2015. « Metabolic bacterial genes and the construction of high-level composite lineages of life ». *Trends in ecology & evolution* 30 (3) : 127-29. <https://dx.doi.org/10.1016%2Fj.tree.2015.01.001>.
- Méheust, Raphaël, Andrew K. Watson, François-Joseph Lapointe, R. Thane Papke, Philippe Lopez et Éric Bapteste. 2018. « Hundreds of novel composite genes and chimeric genes with bacterial origins contributed to haloarchaeal evolution ». *Genome Biol* 19 (1) : 75. <https://doi.org/10.1186/s13059-018-1454-9>.
- Merhej, Vicky et Didier Raoult. 2012. « Rhizome of life, catastrophes, sequence exchanges, gene creations, and giant viruses: how microbial genomics challenges Darwin ». *Frontiers in cellular and infection microbiology* 2 : 113. <https://doi.org/10.3389/fcimb.2012.00113>.
- Mill, John Stuart. 1862. *A System of Logic, Ratiocinative and Inductive Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation*. 4^e éd. Londres : Parker.
- Miller, Alexander. 2019. « Realism ». Dans *The Stanford Encyclopedia of Philosophy*. Éd. hiver 2019. Sous la direction d'Edward N. Zalta. Université de Stanford. <https://plato.stanford.edu/archives/win2019/entries/realism/>.
- Millikan, Ruth Garrett. 1984. *Language, thought and other biological categories*. Cambridge, MA : MIT Press.
- . 2002. « Biofunctions: two paradigms ». Dans *Functions: new readings in the philosophy of psychology and biology*. Sous la direction de Robert Cummins, Andre Ariew et Mark Perlman, 113-43. Oxford : Oxford University Press.

- Millstein, Roberta L. 2006. « Natural selection as a population-level causal process ». *The British Journal for the Philosophy of Science* 57 (4) : 627-53.
<https://doi.org/10.1093/bjps/axl025>.
- Mindell, David P. et Axel Meyer. 2001. « Homology evolving ». *Trends in Ecology & Evolution* 16 (8) : 434-40. [https://doi.org/10.1016/S0169-5347\(01\)02206-6](https://doi.org/10.1016/S0169-5347(01)02206-6).
- Minelli, Alessandro et Giuseppe Fusco. 2013. « Homology ». Dans *The Philosophy of Biology*. Sous la direction de Kostas Kampourakis, 289-322. History, Philosophy and Theory of the Life Sciences. Vol. 1. Dordrecht : Springer. https://doi.org/10.1007/978-94-007-6537-5_15.
- Mooi, Randall D. et Anthony C. Gill. 2010. « Phylogenies without Synapomorphies—A Crisis in Fish Systematics: Time to Show Some Character ». *Zootaxa* 2450 (1) : 26.
<https://doi.org/10.11646/zootaxa.2450.1.2>.
- Morange, Michel. 2017. *Une histoire de la biologie*. Paris : Points.
- Morrison, David A. 2005. « Networks in Phylogenetic Analysis: New Tools for Population Biology ». *International Journal for Parasitology* 35 (5) : 567-82.
<https://doi.org/10.1016/j.ijpara.2005.02.007>.
- . 2010. « Phylogenetic networks in systematic biology (and elsewhere) ». Dans *Research advances in systematic biology*. Sous la direction de Matthen Mohan, 1-48. India : Trivandrum.
- . 2014a. « Phylogenetic Networks: A Review of Methods to Display Evolutionary History ». *Annual Research & Review in Biology* 4 (10) : 1518-43.
<http://dx.doi.org/10.9734/ARRB/2014/8230>.
- . 2014b. « Is the Tree of Life the Best Metaphor, Model, or Heuristic for Phylogenetics? » *Systematic Biology* 63 (4) : 628-38. <https://doi.org/10.1093/sysbio/syu026>.
- . 2015. « Is Sequence Alignment an Art or a Science? » *Systematic Botany* 40 (1) : 14-26.
<https://doi.org/10.1600/036364415X686305>.
- Morrison, David A., Matthew J. Morgan et Scot A. Kelchner. 2015. « Molecular Homology and Multiple-Sequence Alignment: An Analysis of Concepts and Practice ». *Australian Systematic Botany* 28 (1) : 46. <https://doi.org/10.1071/SB15001>.
- Morrison, Margaret C. 1998. « Modelling nature: Between physics and the physical world ». *Philosophia Naturalis* 35 (1), 65-85.
- . 2009. « Models, Measurement and Computer Simulation: The Changing Face of Experimentation ». *Philosophical Studies* 143 (1) : 33-57.
<https://doi.org/10.1007/s11098-008-9317-y>.
- Morrison, Margaret et Mary S. Morgan. 1999. « Models as mediating instruments ». Dans *Models as Mediators*. Sous la direction de Mary S. Morgan et Margaret Morrison, 10-37. Cambridge : Cambridge University Press.
<https://doi.org/10.1017/CBO9780511660108.003>.
- Müller, Gerd B. 2003. « Homology: the evolution of morphological organization ». Dans *Origination of Organismal Form: beyond Gene in Developmental and Evolutionary Biology*. Sous la direction de Gerd B. Müller et Stuart A. Newman, 324-45. Cambridge, MA : MIT Press.
- Mundy, B. 1987. « The Metaphysics of Quantity ». *Philosophical Studies* 51 (1) : 29-54.
<https://doi.org/10.1007/BF00353961>.

- Nakhleh, Luay, Derek Ruths et Li-San Wang. 2005. « RIATA-HGT: A Fast and Accurate Heuristic for Reconstructing Horizontal Gene Transfer ». Dans *Computing and Combinatorics. Lecture Notes in Computer Science*, vol. 3595. Sous la direction de Lusheng Wang, 84-93. Berlin : Springer. https://doi.org/10.1007/11533719_11.
- Naor, Adit, Pascal Lapierre, Moshe Mevarech, R. Thane Papke et Uri Gophna. 2012. « Low Species Barriers in Halophilic Archaea and the Formation of Recombinant Hybrids ». *Curr Biol* 22 (15) : 1444-48. <https://doi.org/10.1016/j.cub.2012.05.056>.
- Naylor, Gavin J. P. et Wesley M. Brown. 1998. « Amphioxus Mitochondrial DNA, Chordate Phylogeny, and the Limits of Inference Based on Comparisons of Sequences ». *Systematic Biology* 47 (1) : 61-76. <https://doi.org/10.1080/106351598261030>.
- Neander, Karen. 1995. « Misrepresenting & malfunctioning ». *Philosophical Studies* 79 (2) : 109-41. <https://doi.org/10.1007/BF00989706>.
- . 2002. « Types of traits: the importance of functional homologues ». Dans *Functions: new readings in the philosophy of psychology and biology*. Sous la direction de Robert Cummins, Andre Ariew et Mark Perlman, 390-415. Oxford : Oxford University Press.
- Nelson, Gareth. 1994. « Homology and Systematics ». Dans *Homology: the hierarchical basis of comparative biology*. Sous la direction de Brian K. Hall, 101-49. San Diego : Academic Press.
- Newman, Mark E. J. 2018. *Networks*. 2^e éd. Oxford : Oxford University Press.
- Niewöhner, Jörg et Margaret Lock. 2018. « Situating Local Biologies: Anthropological Perspectives on Environment/Human Entanglements ». *BioSocieties* 13 (4) : 681-97. <https://doi.org/10.1057/s41292-017-0089-5>.
- Nixon, Kevin C. et James M. Carpenter. 2012. « On Homology ». *Cladistics* 28 (2) : 160-69. <https://doi.org/10.1111/j.1096-0031.2011.00371.x>.
- Noviello, Teresa M. R., Antonella Di Liddo, Giovanna M. Ventola, Antonietta Spagnuolo, Salvatore D'Aniello, Michele Ceccarelli et Luigi Cerulo. 2018. « Detection of Long Non-Coding RNA Homology, a Comparative Study on Alignment and Alignment-Free Metrics ». *BMC Bioinformatics* 19 (1) : 407. <https://doi.org/10.1186/s12859-018-2441-6>.
- O'Connor, Cailin. 2017. « Games and Kinds ». *The British Journal for the Philosophy of Science* 70 (3). <https://doi.org/10.1093/bjps/axx027>.
- O'Halloran, Damien. 2014. « A practical guide to phylogenetics for nonexperts ». *JoVE (Journal of Visualized Experiments)* 84 : e50975. <https://dx.doi.org/10.3791/50975>.
- O'Hara, Robert J. 1988. « Homage to Clio, or, Toward an Historical Philosophy for Evolutionary Biology ». *Systematic Zoology* 37 (2) : 142. <https://doi.org/10.2307/2992272>.
- . 1992. « Telling the tree: narrative representation and the study of evolutionary history ». *Biology and Philosophy* 7 (2) : 135-60. <http://dx.doi.org/10.1007/BF00129880>.
- . 1996. « Trees of history in systematics and philology ». *Memorie della Società Italiana di Scienze Naturali e del Museo Civico di Storia Naturale di Milano* 27 (1) : 81-88.
- . 1997b. « Population Thinking and Tree Thinking in Systematics ». *Zoologica Scripta* 26 (4) : 323-29. <https://doi.org/10.1111/j.1463-6409.1997.tb00422.x>.
- Okasha, Samir. 2002. « Darwinian Metaphysics: Species and the Question of Essentialism ». *Synthese* 131 : 191-213. <https://doi.org/10.1023/A:1015731831011>.
- O'Leary, Maureen A., John Gatesy et Michael J. Novacek. 2003. « Are the Dental Data Really at Odds with the Molecular Data? Morphological Evidence for Whale Phylogeny

- (Re)Reexamined ». *Systematic Biology* 52 (6) : 853-64.
<https://doi.org/10.1093/sysbio/52.6.853>.
- O'Malley, Maureen A. 2013. « When Integration Fails: Prokaryote Phylogeny and the Tree of Life ». *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (4) : 551-62.
<https://doi.org/10.1016/j.shpsc.2012.10.003>.
- . 2016. « Histories of Molecules: Reconciling the Past ». *Studies in History and Philosophy of Science Part A* 55 : 69-83. <https://doi.org/10.1016/j.shpsa.2015.09.002>.
- O'Malley, Maureen A. et Eugene V. Koonin. 2011. « How stands the Tree of Life a century and a half after The Origin? » *Biology Direct* 6 (1) : 32. <https://doi.org/10.1186/1745-6150-6-32>.
- Ottenburghs, Jente. 2019. « Multispecies Hybridization in Birds ». *Avian Research* 10 (1) : 20.
<https://doi.org/10.1186/s40657-019-0159-4>.
- Owen, Richard. 1843. *Lectures on the comparative anatomy and physiology of the invertebrate animals : delivered at the Royal College of Surgeons, in 1843 / By Richard Owen ; From notes taken by William White Cooper and revised by Professor Owen*. Londres : Longman, Brown, Green, and Longmans. <https://doi.org/10.5962/bhl.title.6788>.
- Panchen, Alec L. 1992. *Classification, evolution, and the nature of biology*. Cambridge : Cambridge University Press.
- . 1994. « Richard Owen and the Concept of Homology ». Dans *Homology: the hierarchical basis of comparative biology*. Sous la direction de Brian K. Hall, 21-62. San Diego : Academic Press.
- Papale, François. 2021. « Evolution by Means of Natural Selection without Reproduction: Revamping Lewontin's Account ». *Synthese* 198 (11) : 10429-55.
<https://doi.org/10.1007/s11229-020-02729-6>.
- Papale, François, Jordane Saget et Éric Bapteste. 2020. « Networks Consolidate the Core Concepts of Evolution by Natural Selection ». *Trends in Microbiology* 28 (4) : 254-65.
<https://doi.org/10.1016/j.tim.2019.11.006>.
- Parks, Donovan H., Maria Chuvochina, David W. Waite, Christian Rinke, Adam Skarshewski, Pierre-Alain Chaumeil et Philip Hugenholtz. 2018. « A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life ». *Nature Biotechnology* 36 : 996-1004. <https://doi.org/10.1038/nbt.4229>.
- Pasek, Sophie, Jean-Loup Risler et Pierre Brezellec. 2006. « Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins ». *Bioinformatics* 22 (12) : 1418-23. <https://doi.org/10.1093/bioinformatics/btl135>.
- Patané, José S. L., Joaquim Martins et João C. Setubal. 2018. « Phylogenomics ». Dans *Comparative Genomics. Methods in Molecular Biology, vol. 1704*. Sous la direction de João C. Setubal, Jens Stoye et Peter F. Stadler, 103-87. New York : Springer New York.
https://doi.org/10.1007/978-1-4939-7463-4_5.
- Patterson, Colin. 1982. « Morphological Characters and Homology ». Dans *Problems of Phylogenetic Reconstruction*. Sous la direction de Kenneth Allan Joysey and Adrian E. Friday, 21-74. Londres : Academic Press for the Systematics Association.
- . 1987. *Molecules and morphology in evolution: conflict or compromise*. Cambridge : Cambridge University Press.

- Pavlinov, I. Ya. 2012. « The Contemporary Concepts of Homology in Biology: A Theoretical Review ». *Biology Bulletin Reviews* 2 (1) : 36-54. <https://doi.org/10.1134/S2079086412010057>.
- Pearson, Christopher H. 2019. « Are Homologies Really Natural Kinds? » *Biology & Philosophy* 34 (4) : 42. <https://doi.org/10.1007/s10539-019-9696-9>.
- Peirce, Charles S. 1992. *The essential Peirce: selected philosophical writings*. Vol. 1. Bloomington : Indiana University Press.
- Penny, David. 2011. « Darwin's Theory of Descent with Modification, versus the Biblical Tree of Life ». *PLoS Biology* 9 (7) : e1001096. <https://doi.org/10.1371/journal.pbio.1001096>.
- Penny, David, L. R. Foulds et M. D. Hendy. 1982. « Testing the Theory of Evolution by Comparing Phylogenetic Trees Constructed from Five Different Protein Sequences ». *Nature* 297 (5863) : 197-200. <https://doi.org/10.1038/297197a0>.
- Perelman, Polina, Warren E. Johnson, Christian Roos, Hector N. Seuánez, Julie E. Horvath, Miguel A. M. Moreira, Bailey Kessing et al. 2011. « A Molecular Phylogeny of Living Primates ». Édité par Jürgen Brosius. *PLoS Genetics* 7 (3) : e1001342. <https://doi.org/10.1371/journal.pgen.1001342>.
- Pham, Diem-Trang, Shanshan Gao et Vinhthuy Phan. 2017. « An Accurate and Fast Alignment-Free Method for Profiling Microbial Communities ». *Journal of Bioinformatics and Computational Biology* 15 (03) : 1740001. <https://doi.org/10.1142/S0219720017400017>.
- Phemister, Pauline. 1990. « Real Essences in Particular ». *Locke Newsletter* 21 : 27-55.
- de Pinna, Mario C. 1991. « Concepts and Tests of Homology in the Cladistic Paradigm ». *Cladistics* 7 (4) : 367-94. <https://doi.org/10.1111/j.1096-0031.1991.tb00045.x>.
- Poincaré, Henri. 1898. « The Measure of Time ». Dans *The Value of Science*, 26-36. New York : Dover Publications.
- Popper, Karl. 1959. *The Logic of Scientific Discovery*. Nouv. éd. Londres : Taylor and Francis Group.
- Portik, Daniel M. et John J Wiens. 2021. « Do Alignment and Trimming Methods Matter for Phylogenomic (UCE) Analyses? » *Systematic Biology* 70 (3) : 440-62. <https://doi.org/10.1093/sysbio/syaa064>.
- Potochnik, Angela. 2015. « The Diverse Aims of Science ». *Studies in History and Philosophy of Science Part A* 53 : 71-80. <https://doi.org/10.1016/j.shpsa.2015.05.008>.
- Prangishvili, David, Dennis H. Bamford, Patrick Forterre, Jaime Irazo, Eugene V. Koonin et Mart Krupovic. 2017. « The Enigmatic Archaeal Virosphere ». *Nature Reviews Microbiology* 15 (12) : 724-39. <https://doi.org/10.1038/nrmicro.2017.125>.
- Putnam, Hilary. 1973. « Explanation and reference ». Dans *Conceptual change*. Sous la direction de Glenn Pearce et Patrick Maynard, 199-221. Berlin : Springer. https://doi.org/10.1007/978-94-010-2548-5_11
- . 1990. « Is water necessarily H₂O? » Dans *Realism with a Human Face*. Sous la direction de Jame Conant, 54-79. Cambridge, MA : Harvard University Press.
- Quine, Willard Van Orman. 1961. « Two Dogmas of Empiricism ». Dans *From A Logical Point of View*, 20-46. 2^e éd. rev. Cambridge, MA : Harvard University Press.
- . 1971. « Epistemology naturalized ». *Akten des XIV. Internationalen Kongresses für Philosophie* 6 (1-5) : 87-103. <https://doi.org/10.5840/wcp14197168>

- Quinn, Aleta. 2017. « When Is a Cladist Not a Cladist? » *Biology & Philosophy* 32 (4) : 581-98. <https://doi.org/10.1007/s10539-017-9577-z>.
- Ragan, Mark A. 2009. « Trees and networks before and after Darwin ». *Biol Direct* 4 : 43. <https://doi.org/10.1186/1745-6150-4-43>.
- Ragan, Mark A. et Cheong Xin Chan. 2013. « Biological Intuition in Alignment-Free Methods: Response to Posada ». *Journal of Molecular Evolution* 77 (1-2) : 1-2. <https://doi.org/10.1007/s00239-013-9573-0>.
- Ramsey, William. 2021. « Eliminative Materialism ». Dans *The Stanford Encyclopedia of Philosophy*. Sous la direction d'Edward N. Zalta. Université de Stanford. <https://plato.stanford.edu/archives/fall2021/entries/materialism-eliminative/>.
- Raoult, Didier. 2010. « The Post-Darwinist Rhizome of Life ». *The Lancet* 375 (9709) : 104-5. [https://doi.org/10.1016/S0140-6736\(09\)61958-9](https://doi.org/10.1016/S0140-6736(09)61958-9).
- Raoult, Didier et Eugene V. Koonin. 2012. « Microbial Genomics Challenge Darwin ». *Frontiers in Cellular and Infection Microbiology* 2 : 127. <https://doi.org/10.3389/fcimb.2012.00127>.
- Reichenbach, Hans. 1927. *The Philosophy of Space and Time*. New York : Dover Publications.
- Remane, Adolf. 1952. *Die Grundlagen des Natu"rlichen Systems, der vergleichenden Anatomie und der Phylogenetik*. Königsstein : Otto Koeltz.
- Ren, Fengrong, Hiroshi Tanaka et Ziheng Yang. 2009. « A Likelihood Look at the Supermatrix–Supertree Controversy ». *Gene* 441 (1-2) : 119-25. <https://doi.org/10.1016/j.gene.2008.04.002>.
- Riedl, Rupert. 1978. *Order in Living Organisms: A Systems Analysis of Evolution*. New York : Wiley.
- Rieppel, Olivier. 1994. « Homology, topology, and typology: the history of modern debates ». Dans *Homology: the hierarchical basis of comparative biology*. Sous la direction de Brian K. Hall, 63-100. San Diego : Academic Press.
- . 2005a. « The Philosophy of Total Evidence and Its Relevance for Phylogenetic Inference ». *Papéis Avulsos de Zoologia (São Paulo)* 45 (8). <https://doi.org/10.1590/S0031-10492005000800001>.
- . 2005b. « Modules, Kinds, and Homology ». *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* 304B (1) : 18-27. <https://doi.org/10.1002/jez.b.21025>.
- . 2005c. « Monophyly, Paraphyly, and Natural Kinds ». *Biology & Philosophy* 20 (2-3) : 465-87. <https://doi.org/10.1007/s10539-004-0679-z>.
- . 2009. « 'Total Evidence' in Phylogenetic Systematics ». *Biology & Philosophy* 24 (5) : 607-22. <https://doi.org/10.1007/s10539-008-9122-1>.
- . 2010a. « The series, the network, and the tree: changing metaphors of order in nature ». *Biology & Philosophy* 25 (4) : 475-96. <http://dx.doi.org/10.1007/s10539-010-9216-4>.
- . 2010b. « New Essentialism in Biology ». *Philosophy of Science* 77 (5) : 662-73. <https://doi.org/10.1086/656539>.
- . 2015. « Homology: A philosophical and Biological Perspective ». Dans *Handbook of Paleoanthropology*. Sous la direction de Winfried Henke et Ian Tattersall, 295-316. Berlin : Springer.

- . 2016. *Phylogenetic systematics: Haeckel to Hennig*. Species and systematics. Boca Raton : CRC Press (Taylor & Francis Group).
- Rosenberg, Alex et Frédéric Bouchard. 2005. « Matthen and Ariew's obituary for fitness: reports of its death have been greatly exaggerated ». *Biology and Philosophy* 20 (2) : 343-53. <http://dx.doi.org/10.1007/s10539-005-2560-0>.
- Rosenberg, Eugene et Ilana Zilber-Rosenberg. 2018. « The Hologenome Concept of Evolution after 10 Years ». *Microbiome* 6 (1) : 78. <https://doi.org/10.1186/s40168-018-0457-9>.
- Roughgarden, Joan. 2020. « Holobiont Evolution: Mathematical Model with Vertical vs. Horizontal Microbiome Transmission ». *Philosophy, Theory, and Practice in Biology* 12 (2). <https://doi.org/10.3998/ptpbio.16039257.0012.002>.
- Rupke, Nicolaas A. 1993. « Richard Owen's vertebrate archetype ». *Isis* 84 (2) : 231-51. <https://doi.org/10.1086/356461>.
- Russell, Bertrand. 1948. *Human Knowledge: Its Scope and Limits*. Londres : George Allen and Unwin.
- Saitou, Naruya et Masatoshi Nei. 1987. « The neighbor-joining method: a new method for reconstructing phylogenetic trees. » *Molecular biology and evolution* 4 (4) : 406-25. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>.
- Sanderson, Michael J., Andy Purvis et Chris Henze. 1998. « Phylogenetic Supertrees: Assembling the Trees of Life ». *Trends in Ecology & Evolution* 13 (3) : 105-9. [https://doi.org/10.1016/S0169-5347\(97\)01242-1](https://doi.org/10.1016/S0169-5347(97)01242-1).
- Sapp, J. 2009. *The New Foundations of Evolution: On the Tree of Life*. New-York : Oxford University Press.
- Schaffer, Simon. 1992. « Late Victorian Metrology and its Instrumentation: A Manufactory of Ohms ». Dans *Invisible Connections: Instruments, Institutions, and Science*. Sous la direction de Robert Bud et Suzan E. Cozzens, 23-56. Cardiff : SPIE Optical Engineering.
- Sellars, Wilfrid. 1956. *Empiricism and the philosophy of mind*. Cambridge, MA : Harvard University Press.
- Shapin, Steven et Simon Schaffer. 1985. *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton : Princeton University Press.
- Skerratt, Lee, Nick Campbell, Anna Murrell, Shelley Walton, David Kemp et Stephen Barker. 2002. « The Mitochondrial 12S Gene Is a Suitable Marker of Populations of *Sarcoptes Scabiei* from Wombats, Dogs and Humans in Australia ». *Parasitology Research* 88 (4) : 376-79. <https://doi.org/10.1007/s00436-001-0556-5>.
- Slater, Matthew H. 2013. « Cell types as natural kinds ». *Biological Theory* 7 (2) : 170-79. <http://dx.doi.org/10.1007/s13752-012-0084-9>.
- . 2015. « Natural kindness ». *The British Journal for the Philosophy of Science* 66 (2) : 375-411. <https://doi.org/10.1093/bjps/axt033>.
- . 2018. « Anchoring in Ecosystemic Kinds ». *Synthese* 195 (4) : 1487-1508. <https://doi.org/10.1007/s11229-016-1302-y>.
- Sneath, Peter H. A. et Robert R. Sokal. 1973. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. San Francisco : W. H. Freeman.
- . 1975. « Cladistic Representation of Reticulate Evolution ». *Systematic Zoology* 24 (3) : 360. <https://doi.org/10.2307/2412721>.

- Soames, Scott. 2002. *Beyond rigidity: The unfinished semantic agenda of naming and necessity*. Oxford : Oxford University Press.
- Sober, Elliott. 1980. « Evolution, Population Thinking, and Essentialism ». *Philosophy of Science* 47 (3) : 350-83. <https://doi.org/10.1086/288942>.
- Sokal, Robert R. et Peter H. A. Sneath. 1963. *Numerical taxonomy*. San Francisco : W. H. Freeman.
- Solís-Lemus, Claudia, Paul Bastide et Cécile Ané. 2017. « PhyloNetworks: A Package for Phylogenetic Networks ». *Molecular Biology and Evolution* 34 (12) : 3292-98. <https://doi.org/10.1093/molbev/msx235>.
- Sorokin, Dimitry Y., Kira S. Makarova, Ben Abbas, Manuel Ferrer, Peter N. Golyshin, Erwin A. Galinski, Sergio Ciordia et al. 2017. « Discovery of Extremely Halophilic, Methyl-Reducing Euryarchaea Provides Insights into the Evolutionary Origin of Methanogenesis ». *Nature Microbiology* 2 (8) : 170-81. <https://doi.org/10.1038/nmicrobiol.2017.81>.
- Sorokin, Dimitry Y., Kira S. Makarova, Ben Abbas, Manuel Ferrer, Peter N. Golyshin, Erwin A. Galinski, Sergio Ciordia et al. 2019. « Reply to ‘Evolutionary Placement of Methanonatronarchaeia’ ». *Nature Microbiology* 4 (4) : 560-61. <https://doi.org/10.1038/s41564-019-0358-0>.
- Sorokin, Dimitry Y., Alexander Y. Merkel, Ben Abbas, Kira S. Makarova, W. Irene C. Rijpstra, M. Koenen, Jaap S. Sinninghe Damsté, Erwin A. Galinski, Eugene V. Koonin et Mark C. M. van Loosdrecht. 2018. « Methanonatronarchaeum Thermophilum Gen. Nov., Sp. Nov. and “Candidatus Methanohalarchaeum Thermophilum”, Extremely Halo(Natrono)Philic Methyl-Reducing Methanogens from Hypersaline Lakes Comprising a New Euryarchaeal Class Methanonatronarchaeia Classis Nov. ». *International Journal of Systematic and Evolutionary Microbiology* 68 (7) : 2199-2208. <https://doi.org/10.1099/ijsem.0.002810>.
- de Souza, Sandro J. 2012. « Domain shuffling and the increasing complexity of biological networks ». *Bioessays* 34 (8) : 655-57. <https://doi.org/10.1002/bies.201200006>.
- Spencer, Quayshawn. 2016. « Genuine Kinds and Scientific Reality ». Dans *Natural kinds and classification in scientific practice*. Sous la direction de Catherine Kendig, 157-72. New York : Routledge.
- Springer, Mark S., Robert W. Meredith, John Gatesy, Christopher A. Emerling, Jong Park, Daniel L. Rabosky, Tanja Stadler et al. 2012. « Macroevolutionary Dynamics and Historical Biogeography of Primate Diversification Inferred from a Species Supermatrix ». *PLoS ONE* 7 (11) : e49521. <https://doi.org/10.1371/journal.pone.0049521>.
- Stevens, Peter F. 1984. « Metaphors and Typology in the Development of Botanical Systematics 1690-1960, or the Art of Putting New Wine in Old Bottles ». *Taxon* 33 (2) : 169-211. <https://doi.org/10.2307/1221161>.
- . 1994. *The development of biological systematics: Antoine-Laurent de Jussieu, nature, and the natural system*. New York : Columbia University Press.
- . 1997. « Mind, memory and history: How classifications are shaped by and through time, and some consequences ». *Zoologica Scripta* 26 (4) : 293-302. <https://doi.org/10.1111/j.1463-6409.1997.tb00419.x>.
- Strickland, Hugh Edwin. 1846. « On the structural relations of organized beings ». *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 28 (188) : 354-64. <https://doi.org/10.1080/14786444608645434>.

- Stuart, Matthew. 1999. « Locke on Natural Kinds ». *History of Philosophy Quarterly* 16 (3) : 277-96.
- Suárez, Javier. 2020. « The Stability of Traits Conception of the Hologenome: An Evolutionary Account of Holobiont Individuality ». *History and Philosophy of the Life Sciences* 42 (1) : 11. <https://doi.org/10.1007/s40656-020-00305-2>.
- Suárez, Javier et Vanessa Triviño. 2020. « What Is a Hologenomic Adaptation? Emergent Individuality and Inter-Identity in Multispecies Systems ». *Frontiers in Psychology* 11 : 187. <https://doi.org/10.3389/fpsyg.2020.00187>.
- Suárez-Díaz, Edna. 2014. « The Long and Winding Road of Molecular Data in Phylogenetic Analysis ». *Journal of the History of Biology* 47 (3) : 443-78. <https://doi.org/10.1007/s10739-013-9373-9>.
- Suárez-Gonzalez, Adriana, Christian Lexer et Quentin C. B. Cronk. 2018. « Adaptive Introgression: A Plant Perspective ». *Biology Letters* 14 (3) : 20170688. <https://doi.org/10.1098/rsbl.2017.0688>.
- Suppes, Patrick, David H. Krantz, R. Duncan Luce et Amos Tversky. 1989. *Foundations of Measurement*. Vol. 2. San Diego : Academic Press.
- Swoyer, Chris 1987. « The Metaphysics of Measurement ». Dans *Measurement, Realism and Objectivity*. Sous la direction de John Forge, 235-90. Dordrecht : Reidel.
- Szathmáry, Eörs. 2015. « Toward Major Evolutionary Transitions Theory 2.0 ». *Proceedings of the National Academy of Sciences* 112 (33) : 10104-11. <https://doi.org/10.1073/pnas.1421398112>.
- Tal, Eran. 2012. « The Epistemology of Measurement: A Model-Based Account ». Thèse de doctorat, Université de Toronto. [https://tspace.library.utoronto.ca/bitstream/1807/34936/1/Tal Eran 201211 PhD thesis.pdf](https://tspace.library.utoronto.ca/bitstream/1807/34936/1/Tal%20Eran%20201211%20PhD%20thesis.pdf).
- . 2013. « Old and New Problems in Philosophy of Measurement ». *Philosophy Compass* 8 (12) : 1159-73. <https://doi.org/10.1111/phc3.12089>.
- . 2015. « Measurement in science ». Dans *The Stanford Encyclopedia of Philosophy*. Sous la direction d'Edward N. Zalta. Université de Stanford. <https://plato.stanford.edu/entries/measurement-science/>.
- . 2016. « Making Time: A Study in the Epistemology of Measurement ». *British journal for the philosophy of science* 67 (1) : 297-335. <https://doi.org/10.1093/bjps/axu037>.
- . 2017. « A Model-Based Epistemology of Measurement ». Dans *Reasoning in measurement*. Sous la direction de Nicola Mössner et Alfred Nordmann, 233-53. New York : Routledge.
- Taxis, Tasia M., Sara Wolff, Sarah J. Gregg, Nicholas O. Minton, Chiqian Zhang, Jingjing Dai, Robert D. Schnabel et al. 2015. « The Players May Change but the Game Remains: Network Analyses of Ruminal Microbiomes Suggest Taxonomic Differences Mask Functional Similarity ». *Nucleic Acids Research* 43 (20) : 9600-9612. <https://doi.org/10.1093/nar/gkv973>.
- Than, Cuong, Derek Ruths et Luay Nakhleh. 2008. « PhyloNet: A Software Package for Analyzing and Reconstructing Reticulate Evolutionary Relationships ». *BMC Bioinformatics* 9 (1) : 322. <https://doi.org/10.1186/1471-2105-9-322>.

- Thornton, Stephen. 2021. « Karl Popper ». Dans *The Stanford Encyclopedia of Philosophy*. Éd. automne 2021. Sous la direction d'Edward N. Zalta. Université de Stanford. <https://plato.stanford.edu/entries/popper/>.
- Toleman, M. A., J. Spencer, L. Jones et Timothy Rutland Walsh. 2012. « blaNDM-1 is a chimera likely constructed in *Acinetobacter baumannii* ». *Antimicrobial agents and chemotherapy* 56 (5) : 2773-76. <https://doi.org/10.1128/AAC.06297-11>.
- Tort, Patrick. 1983. *La pensée hiérarchique et l'évolution : les complexes discursifs*. Paris : Aubier Montaigne.
- . 1989. *La raison classificatoire : quinze études*. Paris : Aubier.
- van Fraassen, Bastian C. 1980. *The Scientific Image (Clarendon Library of Logic and Philosophy)*. New York: Clarendon Press.
- . 2008. *Scientific Representation: Paradoxes of Perspective*. Oxford : Oxford University Press.
- . 2012. « Modeling and measurement: The criterion of empirical grounding ». *Philosophy of Science* 79 (5) : 773-84. <https://doi.org/10.1086/667847>.
- van Valen, Leigh M. 1982. « Homology and causes ». *Journal of Morphology* 173 (3) : 305-12. <https://doi.org/10.1002/jmor.1051730307>.
- Vanderpool, Dan, Bui Quang Minh, Robert Lanfear, Daniel Hughes, Shwetha Murali, R. Alan Harris, Muthuswamy Raveendran et al. 2020. « Primate Phylogenomics Uncovers Multiple Rapid Radiations and Ancient Interspecific Introgression ». *PLOS Biology* 18 (12) : e3000954. <https://doi.org/10.1371/journal.pbio.3000954>.
- Velasco, Joel D. 2012. « The Future of Systematics: Tree Thinking without the Tree ». *Philosophy of Science* 79 (5) : 624-36. <https://doi.org/10.1086/667878>.
- . 2013. « The Tree of Life ». Dans *The Cambridge Encyclopedia of Darwin and Evolutionary Thought*. Sous la direction de Michael Ruse, 430-35. Cambridge : Cambridge University Press.
- Verboon, Annemieke R. 2014. « The Medieval Tree of Porphyry: An Organic Structure of Logic ». Dans *The Tree*. Sous la direction de Pippa Saloni et Andrea Worm, 95-116. Turnhout : Brepols Publishers. <https://doi.org/10.1484/M.IMR-EB.1.102023>.
- Vinga, Susana. 2014. « Alignment-free methods in computational biology ». *Briefings in Bioinformatics* 15 (3) : 341-42. <https://doi.org/10.1093/bib/bbu005>.
- Vinga, Susana et Jonas Almeida. 2003. « Alignment-free sequence comparison—a review ». *Bioinformatics* 19 (4) : 513-23. <https://doi.org/10.1093/bioinformatics/btg005>.
- Wägele, Johann et Christoph Mayer. 2007. « Visualizing differences in phylogenetic information content of alignments and distinction of three classes of long-branch effects ». *BMC Evolutionary Biology* 7 (1) : 147. <https://doi.org/10.1186/1471-2148-7-147>.
- Wagner, Günter P. 1989. « The biological homology concept ». *Annual Review of Ecology and Systematics* 20 (1) : 51-69. <https://doi.org/10.1146/annurev.es.20.110189.000411>.
- . 1994. « Homology and the Mechanisms of Development ». Dans *Homology: the hierarchical basis of comparative biology*. Sous la direction de Brian K. Hall, 273-99. San Diego : Academic Press.
- . 1996. « Homologues, Natural Kinds and the Evolution of Modularity ». *American Zoologist* 36 (1) : 36-43. <https://doi.org/10.1093/icb/36.1.36>.

- Watrous, Larry E. et Quentin D. Wheeler. 1981. « The Out-Group Comparison Method of Character Analysis ». *Systematic Biology* 30 (1) : 1-11.
<https://doi.org/10.1093/sysbio/30.1.1>.
- Watson, Andrew K., Michel Habib et Éric Bapteste. 2020. « Phylosystemics: Merging Phylogenomics, Systems Biology, and Ecology to Study Evolution ». *Trends in Microbiology* 28 (3) : 176-90. <https://doi.org/10.1016/j.tim.2019.10.011>.
- Watson, Andrew K., Romain Lannes, Jananan S. Pathmanathan, Raphaël Méheust, Slim Karkar, Philippe Colson, Eduardo Corel, Philippe Lopez et Eric Bapteste. 2019. « The Methodology Behind Network Thinking: Graphs to Analyze Microbial Complexity and Evolution ». Dans *Evolutionary Genomics. Methods in Molecular Biology, vol. 1910*. Sous la direction de Maria Anisimova, 271-308. New York : Springer New York.
https://doi.org/10.1007/978-1-4939-9074-0_9.
- Watson, Andrew K., Philippe Lopez et Eric Bapteste. 2021. « Retracing Lineage History: Time to Emphasize Genetic Turnover ». *Trends in Microbiology* 29 (11) : 957-58.
<https://doi.org/10.1016/j.tim.2021.08.001>.
- Wen, Dingqiao, Yun Yu et Luay Nakhleh. 2017. « Bayesian Inference of Reticulate Phylogenies under the Multispecies Network Coalescent ». *PLOS Genetics* 13 (2) : e1006598.
<https://doi.org/10.1371/journal.pgen.1006598>.
- Wen, Dingqiao, Yun Yu, Jiafan Zhu et Luay Nakhleh. 2018. « Inferring phylogenetic networks using PhyloNet ». *Systematic biology* 67 (4) : 735-40.
<https://doi.org/10.1093/sysbio/syy015>.
- West-Eberhard, Mary Jane. 2003. *Developmental Plasticity and Evolution*. Oxford : Oxford University Press.
- Whewell, William. 1840. *The Philosophy of the Inductive Sciences, Founded Upon Their History*. Cambridge : Cambridge University Press.
- Whiteley, Marvin, Stephen P. Diggle et E. Peter Greenberg. 2017. « Progress in and Promise of Bacterial Quorum Sensing Research ». *Nature* 551 (7680) : 313-20.
<https://doi.org/10.1038/nature24624>.
- Wikforss, Asa Maria. 2005. « Naming Natural Kinds ». *Synthese* 145 (1) : 65-87.
<https://doi.org/10.1007/s11229-004-7013-9>.
- Wiley, Edward O. 1981. *Phylogenetics: The Theory and Practice of Phylogenetic Systematics*. New York : Wiley.
- Wiley, Edward O. et Bruce S. Lieberman. 2011. *Phylogenetics: theory and practice of phylogenetic systematics, second edition*. Hoboken : Wiley.
<https://doi.org/10.1002/9781118017883>.
- Wilkerson, Terence Edward. 1988. « Natural kinds ». *Philosophy* 63 (243) : 29-42.
<https://doi.org/10.1017/S0031819100043114>.
- . 1993. « Species, essences and the names of natural kinds ». *The Philosophical Quarterly* 43 (170) : 1-19. <https://doi.org/10.2307/2219938>.
- Wilkins, John S. 2009. *Species: A history of the idea*. Berkeley : University of California Press.
- Wilkins, John S. et Malte C. Ebach. 2014. *The Nature of Classification: Relationships and Kinds in the Natural Sciences*. Basingstoke : Palgrave Macmillan.

- Williams, David M. 2004. « Homologues and Homology, Phenetics and Cladistics: 150 Years of Progress ». Dans *Milestones in Systematics*. Sous la direction de David M. Williams et Peter L. Forey, 191-224. Boca Raton : CRC Press (Taylor & Francis Group).
- Williams, David M. et Malte C. Ebach. 2012. « Confusing Homologs as Homologies: A Reply to “On Homology”: Letter to the Editor ». *Cladistics* 28 (3) : 223-24. <https://doi.org/10.1111/j.1096-0031.2011.00387.x>.
- Willmann, Rainer. 2003. « From Haeckel to Hennig: The Early Development of Phylogenetics in German-Speaking Europe ». *Cladistics* 19 (6) : 449-79. <https://doi.org/10.1111/j.1096-0031.2003.tb00383.x>.
- Wilson, Robert A. 1996. « Promiscuous Realism ». *The British Journal for the Philosophy of Science* 47 (2) : 303-16. <https://doi.org/10.1093/bjps/47.2.303>.
- Wilson, Robert A., Matthew J. Barker et Ingo Brigandt. 2007. « When Traditional Essentialism Fails: Biological Natural Kinds ». *Philosophical Topics* 35 (1/2) : 189-215. <https://doi.org/10.5840/philtopics2007351/29>.
- Winsberg, Eric. 2012. « Values and uncertainties in the predictions of global climate models ». *Kennedy Institute of Ethics Journal* 22 (2) : 111-37. <https://doi.org/10.1353/ken.2012.0008>.
- Winsor, Mary P. 2006. « Linnaeus’s Biology was not Essentialist ». *Annals of the Missouri botanical Garden* 93 (1) : 2-7. [https://doi.org/10.3417/0026-6493\(2006\)93\[2:LBWNE\]2.0.CO;2](https://doi.org/10.3417/0026-6493(2006)93[2:LBWNE]2.0.CO;2).
- Winther, Rasmus Grønfeldt. 2012. « Interweaving Categories: Styles, Paradigms, and Models ». *Studies in History and Philosophy of Science* 43 (4) : 628-39. <https://doi.org/10.1016/j.shpsa.2012.07.005>.
- Witteveen, Joeri. 2016. « “A Temporary Oversimplification”: Mayr, Simpson, Dobzhansky, and the Origins of the Typology/Population Dichotomy (Part 2 of 2) ». *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 57 : 96-105. <https://doi.org/10.1016/j.shpsc.2015.09.006>.
- . 2020. « Linnaeus, the Essentialism Story, and the Question of Types ». *TAXON* 69 (6) : 1141-49. <https://doi.org/10.1002/tax.12346>.
- Wittgenstein, Ludwig. 1969. *On certainty*. Traduit par Denis Paul et Gertrude Elizabeth Margaret Anscombe. Sous la direction de Gertrude Elizabeth Margaret Anscombe et Georg Henrik Von Wright. Oxford : Blackwell.
- Woese, Carl R. 1994. « There Must Be a Prokaryote Somewhere: Microbiology’s Search for Itself ». *Microbiological Reviews* 58 (1) : 1-9. <https://doi.org/10.1128/mr.58.1.1-9.1994>.
- . 1987. « Bacterial Evolution ». *Microbiological Reviews* 51 (2) : 221-71. <https://doi.org/10.1128/mr.51.2.221-271.1987>.
- . 1998. « Default Taxonomy: Ernst Mayr’s View of the Microbial World ». *Proceedings of the National Academy of Sciences* 95 (19) : 11043-46. <https://doi.org/10.1073/pnas.95.19.11043>.
- . 2002. « On the evolution of cells ». *Proceedings of the National Academy of Sciences* 99 (13) : 8742-47. <https://doi.org/10.1073/pnas.132266999>.
- Woese, Carl R., Otto Kandler et Mark L. Wheelis. 1990. « Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya. » *Proceedings of*

- the National Academy of Sciences* 87 (12) : 4576-79.
<https://doi.org/10.1073/pnas.87.12.4576>.
- Yu, Yun, R. Matthew Barnett et Luay Nakhleh. 2013. « Parsimonious Inference of Hybridization in the Presence of Incomplete Lineage Sorting ». *Systematic Biology* 62 (5) : 738-51.
<https://doi.org/10.1093/sysbio/syt037>.
- Yu, Yun, Jianrong Dong, Kevin J. Liu et Luay Nakhleh. 2014. « Maximum Likelihood Inference of Reticulate Evolutionary Histories ». *Proceedings of the National Academy of Sciences* 111 (46) : 16448-53. <https://doi.org/10.1073/pnas.1407950111>.
- Yutin, Natalya, Pere Puigbò, Eugene V. Koonin et Yuri I. Wolf. 2012. « Phylogenomics of Prokaryotic Ribosomal Proteins ». *PLoS ONE* 7 (5) : e36972.
<https://doi.org/10.1371/journal.pone.0036972>.
- Zachos, Frank E. 2016. *Species Concepts in Biology*. Cham : Springer.
<https://doi.org/10.1007/978-3-319-44966-1>.
- Zamani-Dahaj, Seyed Alireza, Mohamed Okasha, Jakub Kosakowski et Paul G. Higgs. 2016. « Estimating the Frequency of Horizontal Gene Transfer Using Phylogenetic Models of Gene Gain and Loss ». *Molecular Biology and Evolution* 33 (7) : 1843-57.
<https://doi.org/10.1093/molbev/msw062>.
- Zhang, Chi, Huw A Ogilvie, Alexei J Drummond et Tanja Stadler. 2018. « Bayesian Inference of Species Networks from Multilocus Sequence Data ». *Molecular Biology and Evolution* 35 (2) : 504-17. <https://doi.org/10.1093/molbev/msx307>.
- Zielezinski, Andrzej, Hani Z. Girgis, Guillaume Bernard, Chris-Andre Leimeister, Kujin Tang, Thomas Dencker, Anna Katharina Lau et al. 2019. « Benchmarking of Alignment-Free Sequence Comparison Methods ». *Genome Biology* 20 (1) : 144.
<https://doi.org/10.1186/s13059-019-1755-7>.
- Zielezinski, Andrzej, Susana Vinga, Jonas Almeida et Wojciech M. Karlowski. 2017. « Alignment-Free Sequence Comparison: Benefits, Applications, and Tools ». *Genome Biology* 18 (1) : 186. <https://doi.org/10.1186/s13059-017-1319-7>.
- Zuckerkandl, Emile. 1987. « On the Molecular Evolutionary Clock ». *Journal of Molecular Evolution* 26 (1-2) : 34-46. <https://doi.org/10.1007/BF02111280>.