Université de Montréal

Laisser sa trace : utiliser les interactions pour comprendre l'évolution

par

Mathilde Besson

Département de Sciences Biologiques

Faculté des Arts et Sciences

Thèse présentée en vue de l'obtention du grade de

Philosophiæ Doctor (Ph.D.)

en Biologie

31 décembre 2021

Université de Montréal

Faculté des études supérieures et postdoctorales

Cette thèse intitulée

Laisser sa trace : utiliser les interactions pour comprendre l'évolution

présentée par

Mathilde Besson

a été évaluée par un jury composée des personnes suivantes :

Sandra Binning
(présidente-rapportrice)

Timothée Poisot
(directeur de recherche)

Dominique Gravel
(co-superviseur de recherche)

Jesse Shapiro
(membre du jury)

Élisa Thébault
(examinatrice externe)

Julie Talbot
(représentante du doyen de la FESP)

Résumé

Les interactions font partie intégrante des écosystèmes. Que ce soit aux niveaux les plus fins, comme les protéines, ou les plus larges, comme les méta-communautés, il est possible de les regrouper en réseaux et d'en étudier la structure. Cela a permis de mettre en évidence que certaines structures sont observables à différents niveaux, c'est le cas par exemple des réseaux emboîtés. De plus, les réseaux d'interaction ont la spécificité de ne pas être fixes dans le temps et l'espace, ce qui leur confère un avantage de taille pour l'étude de l'évolution. Ils peuvent ainsi servir de support à l'études des mécanismes intervenants dans les processus évolutifs. Cependant, il n'existe pas encore de méthodologie ayant fait consensus sur l'utilisation des réseaux et leur analyse à différentes échelles d'organisation.

Cette thèse se base sur l'hypothèse que les réseaux, de par leurs propriétés, sont pertinents à considérer pour comprendre l'évolution et ce à différentes échelles d'organisation, et offrent la possibilité de faire des liens entre chacune d'entre elles. L'approche basée sur les réseaux, combinée à l'utilisation de modèles théorique serait donc un outil méthodologique puissant dans l'élargissement des connaissances concernant les processus sous-jacents à l'évolution.

La thèse qui suit composée de six chapitres dont le contenu est le suivant. Elle commence par un chapitre d'introduction aux concepts d'intérêts, notamment sur l'évolution et la coévolution. Le deuxième chapitre est une introduction à l'utilisation des réseaux en écologie, suivit par le troisième chapitre qui effectue une revue non exhaustive des méthodologies développées autour des réseaux d'interactions. Les chapitres suivants sont en quelque sorte une mise en pratique de ces méthodes et ce à différents niveaux d'organisation. Le quatrième chapitre revient sur une étape avortée de ce doctorat qui servira tout de même à la construction du modèle du chapitre suivant. Le cinquième chapitre se concentre sur la coévolution et son suivit au travers des réseaux d'interaction entre les bactéries et leurs virus. Enfin, le sixième chapitre traque l'évolution des communautés grâce à la structure des arbres phylogénétiques et structure des réseaux d'interactions au cours du temps.

Mots clefs : Réseaux d'interactions, Macro-évolution, Co-évolution, Écologie computationnelle, Communautés, Bactéries, Phages, Modèle mathématique, Évolution, Structure des réseaux

Abstract

Interactions are an integral part of ecosystems. Whether at the finest levels, such as proteins, or the broadest, such as meta-communities, it is possible to group them into networks and study their structure. This made it possible to demonstrate that certain structures can be observed at different levels, such as nested networks, for example. In addition, interaction networks have the property of not being fixed in time and space, which gives them a major advantage for the study of evolution. They can thus serve as a support for the study of the mechanisms involved in the evolutionary processes. However, there is not yet a methodology that has achieved consensus on the use of networks and their analysis at different organizational scales.

This thesis is based on the hypothesis that networks, by virtue of their properties, are relevant to consider in order to understand evolution at different organizational scales, and offer the possibility of making links between each of them. The network-based approach, combined with the use of theoretical models, would therefore be a powerful methodological tool in expanding knowledge about the processes underlying evolution.

The thesis which follows consists of six chapters whose content is as follows. It begins with an introductory chapter to the concepts of interest, in particular on evolution and coevolution. The second chapter is an introduction to the use of networks in ecology, followed by the third chapter which performs a non-exhaustive review of the methodologies developed around interaction networks. The following chapters are in a way a practical application of these methods at different levels of organization. The fourth chapter returns to an aborted stage of this doctorate which will nevertheless be used to construct the model of the following chapter. The fifth chapter focuses on coevolution and its follow-up through the interaction networks between bacteria and their viruses. Finally, the sixth chapter tracks the evolution of communities thanks to the structure of phylogenetic trees and the structure of interaction networks over time.

Table des matières

Liste des tableaux

Liste des figures

Acronymes

- ARD: Arm race dynamic - Dynamique de course à l'armement

- FSD: Fluctuating selection dynamic - Dynamique de sélections fluctuante

- GFG: Gene for Gene model - modèle gène pour gène

- MA: Matching Allele model - modèle d'allèle correspondant

- GMCT: Geographic Mosaic Coevolution Theory - Théorie de la mosaïque géographique de la coévolution

Dédicace

"C'qui compte c'est pas l'arrivée, c'est la quête"

- Orelsan, La Quête

Remerciements

Bon, ça va être un peu long, mais c'est que j'ai eu la chance de croiser tellement de personnes "avec une belle énergie" au cours de ce doctorat que celles que je remercie ici n'en sont qu'un échantillon, un échantillon biaisé bien sûr. Je te laisse le loisir de tout lire ou d'aller chercher seulement les remerciements qui t'intéressent !

Tout l'abord, je tiens à remercier les membres du jury, Sandra A. Binning, Elisa Thébault et Jesse Shapiro. Merci pour votre temps et vos commentaires pertinents et constructifs, ils m'ont permis de bonifier cette thèse. Un grand merci spécifiquement à Sandra, pour ta bienveillance et ta compréhension dans le processus.

Ensuite, et comme il est de coutume, je remercie bien évidemment mes directeurs Timothée et Dominique. À tous les deux, merci pour votre temps et votre patience. Merci de m'avoir accompagné durant ces années et de m'avoir offert toutes ces opportunités de rencontre et d'échanges qui m'ont permis d'avancer. Timothée, une sage personne m'a dit un jour que devenir mentor, c'est comme faire des crêpes, au début cela peut paraître un peu chaotique, mais avec le temps, on prend le coup de main. Je ne suis qu'en partie d'accord avec lui, dans les crêpes il n'y a pas juste la personne qui les fait à prendre en compte, mais également la pâte, la poêle et la plaque. Pour la poêle et la plaque, on fait avec ce qu'on a, et parfois, la magie opère. Par contre, pour la pâte, c'est une autre histoire et il est parfois difficile de faire des crêpes avec une pâte à gâteau. Merci d'avoir essayé. Tu as cette capacité de composer un laboratoire où règne la bienveillance et l'ouverture et c'est ce qui fait que j'ai pu trouver mon nouveau projet.

Donc à tous les deux, entité "Directeurs", Merci pour tout ce que j'ai pu apprendre grâce à vous.

À tous les co-auteurs·trices et personnes dont j'ai croisé la route, merci ! Je vous cite, sans ordre particulier, mais vos échanges ont toujours été très enrichissants, chacun à leur manière. Merci à : Eva, Marie-Hélène, Laura Burkle, Guilo Dalla Riva, Marie-Josée Fortin, David Hembry, Erica Newman, Jens Olesen, Mathias Pires, Justin Yeakel, Zachary Freedman, Alicia McGrew, Benjamen Baiser, Sydne Record, Lauren Trotta et Nicholas Gotelli. Un grand merci également à Oliver Kaltz, Joshua Weitz, Paul Rainey, Marcus Aguiar, Paulo Guimaraes, Francois-Joseph Lapointe, Flavia et tou·te·s les autres, pour votre temps et les discussions que nous avons pu avoir qui m'ont permis d'avoir de nouvelles idées ou tout simplement de mieux comprendre ce que je faisais.

À partir de maintenant, je mixe un peu le professionnel et le personnel. Il est parfois difficile de faire la part des choses entre les deux durant le doctorat et cela demeure encore aujourd'hui. L'écrit n'est pas mon fort, j'haïe même plutôt ça, je lui préfère de loin les petits gestes et attentions concrètes pour les personnes qui me sont chères. J'espère ne blesser personne si jamais j'oublie certains ou certaines d'entre vous ou que les mots ne sortent pas comme ils le devraient.

Je tiens, et plus que tout, à remercier ma famille, de cœur comme de sang, toujours présente dans les moments importants, qu'ils soient joyeux ou tristes.

Tout d'abord merci à mes parents, parce qu'iels m'ont fait et ont contribué à la personne que je suis aujourd'hui, merci pour votre soutien infaillible, les colis et les gaufres qui ont fait le bonheur des papilles de bien des personnes ! Alix, je pense que je n'ai jamais autant apprécié que tu fasses de la route en voiture, ces instants suspendus me font énormément de bien et me permettent de revoir mes perspectives. Ces six années auraient été bien différentes sans ta présence. Bien sûr, Xavier et Etienne, même si nos échanges sont beaucoup moins réguliers, c'est toujours rassurant de savoir que vous êtes là. Je suis fière de faire partie de cette famille qui me soutient et m'accompagne dans mes choix.

À ma famille d'ici, je pense que ne vous remercierai jamais assez non plus. Will, je ne regretterai jamais de t'avoir amadoué avec des viennoiseries en échange de quoi j'ai eu le droit à ta patience et ton aide. Malgré ce que tu peux penser, tu es un de mes piliers, sans qui il aurait été bien plus difficile de passer au travers de ces 5 dernières années, tu es et as été un moteur non-négligeable dans mes réflexions et choix de vie. Pour tout ce que tu es, merci. Eva, plus qu'une collègue, j'ai rencontré une amie, je pense bien être condamnée à rire à l'ensemble de tes blagues pour le reste de nos vies ! Merci de m'avoir soutenu et ouvert tes portes durant ce doctorat et surtout la pandémie, je n'ai qu'un petit aperçu de comment cela se serait passé sans toi (et Réglisson et Mog ! mais aussi Brieuc qui n'est pas un petit animal !) et je me sens chanceuse de faire partie de tes amies. Julie, l'organisation de BBQ et autres activités pour les cycles sup nous ont un peu servit de Tinder mais pour les amis et j'en suis bien heureuse ! Merci pour nos marches, cafés et discussions, tu me pousses à me dépasser et à diversifier ma culture cinématographique tout en dégustant de délicieux pop-corn préparé avec soin par Mathieu (que je remercie également au passage pour ses partages de vidéo de l'émission Steap-Tease et de capybaras !).

Je ne veux pas oublier non plus Ines, Michou ou encore Becky, qui m'avez aidé, chacun à votre manière à me trouver et m'affirmer.

Chapitre 1

Introduction

Il n'existe aucun organisme vivant qui n'interagisse pas avec un autre. Ces interactions fournissent des services écologiques et sont à la base du monde que nous connaissons aujourd'hui. Bien avant le début de l'humanité, et même aujourd'hui, les espèces en interaction **évoluent** et parfois **coévoluent** (voir *Glossaire*) pour survivre dans un environnement en constant mouvement. Ces évolutions leur permettent d'être **localement adaptées** à l'environnement abiotique et biotique dans lequel elles se développent.

La compréhension de l'évolution et de la coévolution entre organismes est essentielle pour prédire la possible évolution des écosystèmes qui nous entourent. Commencée il y a plus d'un demi-siècle, l'étude de la coévolution a couvert différents niveaux d'organisation, allant des interactions entre individus, en passant par les populations et jusqu'aux méta-populations.

Les réseaux d'interactions entre espèces s'avèrent être un outil efficace pour étudier la coévolution entre populations (Weitz et al. 2013). Suivant un cadre emprunté à la **théorie des graphes**, les communautés écologiques peuvent être représentées comme un *graphe* où les espèces sont représentées par des *nœuds* et les interactions entre elles par des *liens*. Le graphe est dit dirigé si les arêtes donnent des informations sur la direction de l'interaction (*e.g.* l'espèce 1 mange l'espèce 2), tandis qu'il est dit non dirigé lorsqu'il n'y a aucune information sur la directionnalité des liens. Si les arêtes peuvent être définies quantitativement, le graphe est défini comme pondéré et non pondéré dans le cas contraire. La figure 1.1 illustre différentes combinaisons de propriétés de graphes. Il est courant de voir le terme *réseau* au lieu de graphe. Les projections de réseaux sont couramment utilisées sous deux points de vue de visualisation : projection unipartite et bipartite (figure 1.1). La projection *unipartite* est une représentation simple des réseaux permettant de voir les connexions entre les nœuds. La projection bipartite classe les nœuds en deux catégories (*par exemple* hôtes et parasites, plantes et pollinisateurs). Il est possible d'ajouter *k* "catégories" au sein du réseau, la projection est alors appelée *k-partite* (*e.g.* hôte/parasite/hyper-parasite).

La représentation mathématique du réseau se fait au moyen d'une *matrice d'adjacence*, où
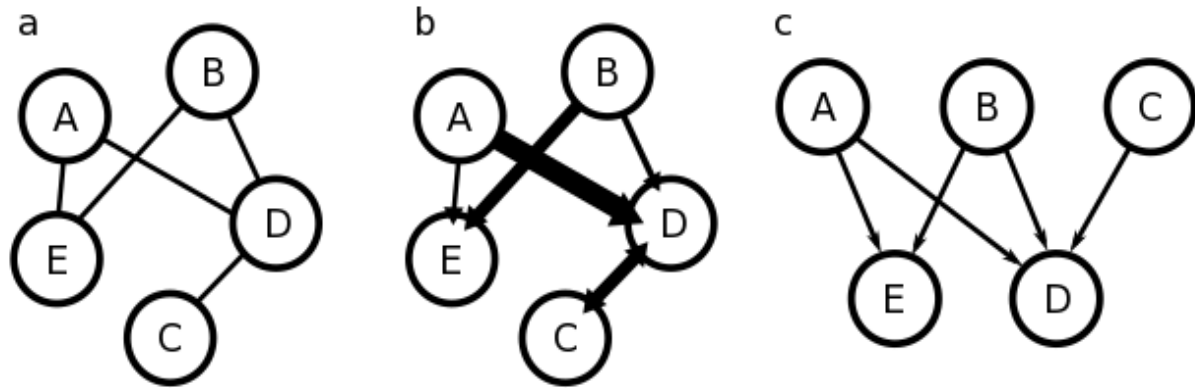
Figure 1.1 Exemple de réseaux d'interaction, où les lettres de A à E représentent différents individus, populations ou espèces. (a) Réseau non directionnel et non pondéré, (b) réseau directionnel et pondéré, (c) réseaux directionnel et non pondéré. (a) et (b) sont des projections de réseaux unipartites et (c) est une projection bipartite avec la distinction spaciale entre deux catégories d'individus/populations/espèces.

chaque interaction entre deux espèces est représentée par un 1, dans un réseau non pondéré, à l'intersection des espèces 1 et 2. La matrice de projection unipartite est une *matrice carrée symétrique*, où la diagonale est généralement remplie de 0 (aucune interaction). Pour une projection bipartite, les dimensions de la matrice dépendent du nombre d'espèces en interaction.

Glossaire

Adaptation locale Une population est localement adaptée à un environnement spécifique lorsqu'elle présente un fitness plus élevé dans cet environnement que dans les autres populations. Elle est influencée par des facteurs génétiques, mais aussi par des facteurs biotiques et abiotiques.

Évolution Changements des traits héréditaires du cycle de vie des organismes survenant sur plusieurs générations. Ces changements peuvent être induits par une adaptation locale.

Coévolution Dynamique évolutive entre deux ou plusieurs organismes ou espèces en interaction. L'influence d'une espèce va induire des changements au fil des générations sur les traits de la seconde espèce. Ce changement induira ensuite des changements dans les traits de l'autre espèce, et ainsi de suite.

Dynamique de course à l'armement ou ARD Dynamique d'adaptation et de contre-adaptation de deux populations ou espèces. Le bras de fer peut être symétrique ou asymétrique. Dans le premier cas, les pressions de sélection sont de même sens pour les deux organismes (*ex.* compétition pour les ressources) ; dans le second cas, il s'agit de pressions de sélection contraires (*ex.* proie et prédateur). Exemple : augmentation de l'infectivité du parasite et de la résistance de l'hôte au fil du temps.

Red Queen Hypothesis ou Reine rouge Hypothèse évolutive, proposée par Van Valen (1973), selon laquelle les populations doivent constamment s'adapter et évoluer en fonction de l'environnement et des espèces en interaction pour éviter l'extinction. Inspirée par *Alice au pays des merveilles*, cette théorie vise à expliquer le taux d'extinction par la co-évolution d'espèces concurrentes et l'avantage de conserver la reproduction sexuée au niveau individuel.

Dynamique de sélection fluctuante ou FSD La dynamique de la sélection naturelle est pilotée par une sélection dépendant de la fréquence dans le temps. Cette dynamique a été proposée pour expliquer le maintien de la diversité génétique. Exemple : un hôte sera plus résistant à ses parasites contemporains qu'aux parasites passés et futurs.

Théorie des graphes Cadre permettant d'étudier la structure mathématique (graphe) qui représente la relation entre les objets (ici, espèces, populations ou individus). Elle a été utilisée pour analyser les réseaux de lignes électriques, les réseaux sociaux et les réseaux écologiques.

Virus et phages Les virus, ou virions, sont des entités biologiques qui peuvent être considérées comme parasites obligatoires puisqu'il leur est nécessaire d'infecter d'autres organismes pour se répliquer. Ils sont constitués d'ADN ou d'ARN encapsulés par une membrane composée de protéines appelées capside. Certains virus sont capables d'infecter des bactéries, ils sont appelés bactériophages ou phages. Les phages présentent deux styles de vie différents : le style lytique, où l'infection induit la mort de la bactérie après l'utilisation de la machinerie de réplication bactérienne, et le style lysogène, où les virus peuvent rester inactifs à l'intérieur de la bactérie pendant plusieurs générations pour entrer, plus tard, dans le style lytique.

1.1 Le cas spécial de la coévolution

L'histoire des études sur l'évolution et la coévolution a commencé avec le questionnement de Charles Darwin, qui se demandait comment les espèces évoluent pour être adaptées localement à leur environnement (Darwin et al. 1859). Il a observé que la prédation et la compétition pouvaient conduire à l'adaptation des espèces en interaction. Darwin n'a cependant pas été le seul à réfléchir à l'idée d'adaptation locale et d'évolution des espèces. Alfred Russel Wallace est arrivé, de façon indépendante, à la conclusion que la sélection naturelle induisait une adaptation. À partir de ce moment-là, l'idée derrière la coévolution, et des possibles mécanismes sous-jacents, commence à émerger et doucement faire son chemin.

Intéressé par les mécanismes génétiques de la résistance aux maladies, Mode (1958) a proposé un modèle mathématique de coévolution. Ce modèle est basé sur le modèle mécaniste de Flor (Flor 1956), avec lequel il a découvert la coévolution gène pour gène (GFG) entre un hôte et son parasite (voir *section 1.3.1.*). Dans ce bouillonnement d'idées et de concepts, Ehrlich & Raven (1964) émerge avec un article fondateur, décrivant la coévolution comme une interaction défense - contre-défense des espèces qui conduit à la co-spéciation et à la co-diversification de ces espèces. Cette vision de la coévolution, également appelée *escape and radiate*, sera plus tard intégrée dans l'hypothèse de la **Red Queen** (Van Valen 1973). A cette époque, aucune définition claire de la coévolution n'était fournie. C'est seulement avec Janzen (1980) que le terme de coévolution obtient une définition claire. Il définit la coévolution comme une évolution réciproque de traits spécifiques d'histoire de vie entre deux populations en interaction.

Toutes ces idées et points de vue ont été des pistes de réflexion clefs sur ce qu'est la coévolution. Ils ont ouvert la voie à plusieurs nouvelles théories de coévolution.

1.2 Les théories de la coévolution

1.2.1 Coévolution par paire et coévolution diffuse

Après qu'une définition claire de la coévolution ait été proposée, plusieurs directions dans l'étude de la coévolution ont été explorées et deux visions opposées de la coévolution ont émergé. D'une part, la coévolution par paire (ou coévolution spécifique), représentant le point de vue de Janzen sur la coévolution, est basée sur l'idée que la coévolution se produit nécessairement et uniquement entre deux populations. Dans ce cas, les changements dans une population induisent des changements adaptatifs directs dans l'autre population. En revanche, la coévolution diffuse (ou coévolution de guilde) stipule que la coévolution multi-spécifique est possible. De ce point de vue, la coévolution dépend du contexte de la communauté (*i.e.* présence, absence, comportement ou abondance d'autres espèces). Au moins l'une des trois conditions suivantes est requise pour favoriser la coévolution diffuse : (i) la sélection réciproque entre deux populations est modifiée par la présence, l'absence ou l'abondance d'une troisième population, (ii) l'appariement génétique avec plusieurs populations, (iii) les "traits d'interaction" de deux populations coévolutives changent à cause d'une troisième population (Fox 1988). Ce point de vue a été extrêmement important pour la compréhension de la structure des communautés (Whitham et al. 2003).

1.2.2 La mosaïque géographique de la coévolution

En même temps, un cadre complémentaire pour expliquer le fait que la coévolution soit parfois différente en fonction de la population émerge. Sur la base d'observations empiriques selon lesquelles une population peut s'adapter et se spécialiser par rapport à une autre population de différentes manières d'une zone géographique à l'autre, Thompson (Thompson 2005) a proposé une description de la coévolution appelée **théorie de la mosaïque géographique de la coévolution** (GMTC).

Selon cette théorie, la coévolution est le résultat d'un mélange de processus génétiques et écologiques. Elle vise à mieux comprendre comment les flux de gènes entre différentes régions conduisent à une dynamique de coévolution entre les populations d'une même région. La GMTC soutient l'hypothèse que la coévolution provient de variations affectant les interactions entre espèces. La GMTC s'appuie sur trois composantes qui induisent ces variations.

La **mosaïque géographique de sélection** est la première. Elle est définie par le fait que les gènes d'une population s'expriment différemment selon l'environnement dans lequel se trouve cette population.

La deuxième composante est la présence de points chauds et froids coévolutifs. En d'autres termes, dans certaines régions, les populations ne contribueront que partiellement, voire pas du tout, au processus de coévolution avec peu ou pas d'influence sur la fitness des populations qu'elles côtoient (*e.g.* relations de commensalisme). Ces régions sont appelées **cold spots** ou **point froid** (figure 1.2). À l'inverse, dans d'autres régions, les populations auront un impact réciproque sur la valeur adaptative des populations avec lesquelles elles interagissent (*e.g.* relations mutuelles ou antagonistes), et sont donc importantes pour le processus de coévolution. Ces régions sont appelées **hot spots** ou **points chauds** (figure 1.2). La présence de points chauds et froids reflète les variations de l'intensité de la sélection réciproque d'une population à l'autre en fonction de l'environnement.

Enfin, la troisième composante du GMTC est caractérisée par le **mélange de caractères** dans les populations. La structure génétique d'une population est en perpétuel changement sous l'effet d'influences internes, comme la fixation des mutations ou la dérive génétique dans les populations, et d'influences externes, comme le flux de gènes entre les populations via la migration ou encore la dynamique d'extinction et de colonisation des populations.
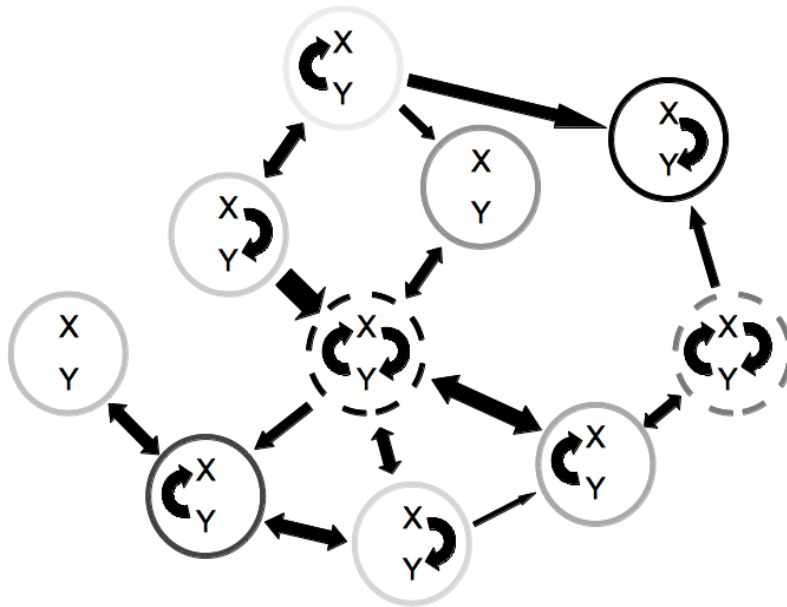
Figure 1.2 Représentation schématique des points chaud et froid. Les cercles représentent des communautés géographiquement distinctes. Les différences abiotiques entre les communautés sont indiquées par la nuance de gris. Les cercles discontinus représentent les points chauds, et les cercles pleins représentent les points froid. Les flèches à l'intérieur des cercles représentent l'effet de la population X sur le fitness de la population Y, et inversement. Les flèches entre les cercles représentent le flux générique entre les communautés et leur épaisseur indique l'intensité de ce flux.

1.2.3 Limitations de ces théories

Même si ces théories ont été cruciales pour comprendre les mécanismes sous-jacents à la coévolution, elles comportent chacune des limites respectives. Les mesures utilisées pour étudier la coévolution par paire ou diffuse, manquent cruellement de preuves empiriques (voir Carmona et al. 2015). Les mesures de coévolution diffuse ne prennent pas en compte les changements évolutifs réels, se concentrant sur la mesure de la coévolution sur une seule génération (Turcotte et al. 2012). Ceci est problématique car, comme le soutient la GMTC, la coévolution pourrait varier dans le temps et l'espace.

De son côté, la GMTC prédit (i) que la coévolution entre deux populations sera fortement influencée par la distribution géographique des points chauds et froids et par le flux génétique entre les populations, et (ii) que la mosaïque de sélection a un rôle décisif dans l'adaptation locale d'une population (Gomulkiewicz et al. 2000 ; Nuismer 2006). Cependant, l'utilisation de la GMTC est rendue difficile par le flou qui existe dans la définition de notions importantes telles que l'appariement ou non des traits entre des populations en interaction, ou la sélection réciproque et la détection de points froids (Gomulkiewicz et al. 2007).

En outre, la GMTC peut conduire à n'importe quel profil d'adaptation ou de maladaptation locale (Nuismer 2006), et les prédictions de coévolution peuvent être obtenues sans tenir compte de la mosaïque de sélection, du mélange de traits ou de la présence de points chauds et froids. Ceci implique que, même si les composantes prises en compte par la GMTC sont importantes pour générer un processus de coévolution entre deux populations, la GMTC ne fait pas de prédictions claires et testables.

En dépit de ces limites, la coévolution diffuse et la GMTC sont complémentaires et nécessaires pour mieux comprendre les processus coévolutifs.

1.3 Comprendre la coévolution

1.3.1 Gene-For-Gene et Matching-Allele modèles

Différents modèles ont été proposés pour explorer la coévolution entre deux populations en interaction (*e.g.* hôte et parasite ou plante et pollinisateur). Basés sur les principes de la génétique des populations, le modèle **Gène-Pour-Gène** (GFG) et le modèle d'**Allèle correspondant** (MA) (Agrawal* & Lively 2002) ont été très importants pour la compréhension des mécanismes sous-jacents de la coévolution. Ces deux modèles reflètent deux extrêmes de la dynamique coévolutive. À titre d'illustration, prenons le cas des interactions hôte-parasite. Le modèle GFG, associé à la **dynamique de course à l'armement** (ARD), est basé sur le lien entre les coûts associés à la résistance de l'hôte et l'infectivité du parasite. Les hôtes et les parasites sont classés selon deux modalités. L'hôte peut être *susceptible* ou *résistant* à l'attaque du parasite. Le parasite peut être *infectieux* ou *non-infectieux* pour son hôte. Dans ce contexte, un hôte sensible peut être infecté par potentiellement tous les parasites, à l'inverse, un hôte résistant ne peut être infecté que par un parasite infectieux. Des mutations fixes dans les populations de parasites ou d'hôtes vont leur permettre d'étendre leurs performances d'infectivité ou de résistance au cours du temps, générant un phénomène de généralisation. Le modèle MA, associé quant à lui à la **dynamique de sélection fluctuante** (FSD), est basé sur le fait que le génotype d'un parasite doit correspondre exactement au génotype de l'hôte pour pouvoir l'infecter. Souvent associé à la théorie de la reine rouge (Van Valen 1973) selon laquelle une population doit continuellement évoluer si elle veut suivre l'évolution des populations avec lesquelles elle interagit, le modèle MA va favoriser la sélection des phénotypes rares de l'hôte, suivie de la sélection du phénotype correspondant du parasite.

1.3.2 Structures emboitée et modulaire

Les prédictions des modèles GFG et MA fournissent une structure génétique spécifique, qui est bien reflétée par la structure des réseaux d'interaction. L'étude de cette structure nous permet de comprendre les processus d'adaptation locale et de co-évolution entre plusieurs populations (Weitz et al. 2013).

Un réseau a une structure **emboîtée** lorsqu'il contient des populations de spécialistes et de généralistes qui peuvent être ordonnées selon leur niveau de généralité (figure 1.3 a) (Atmar & Patterson 1993). Par exemple, un parasite évolue pour augmenter sa capacité à infecter son hôte tout en conservant la capacité d'infecter son hôte *ancestral*. La plupart des études sur la structure des réseaux antagonistes mettent l'accent sur la généralisation des populations au fil du temps, dessinant une structure emboîtée des réseaux (Buckling & Rainey 2002 ; Hall et al. 2011a ; Leggett et al. 2013 ; Scofield et al. 2015).

À l'extrême opposé, un réseau d'interaction peut avoir une structure **modulaire** (figure 1.3 b). Elle est caractérisée par des groupes de populations interagissant davantage entre eux qu'avec d'autres populations, formant des modules ou des compartiments (Bascompte et al. 2003). Ce type de structure provient de l'adaptation des populations aux populations avec lesquelles elles interagissent.

En associant les études de la structure du réseau et de la phylogénie des populations, Cattin et al. (2004a) a mis en évidence le lien entre la structure génétique et la structure du réseau. La structure emboîtée correspond aux prédictions du modèle GFG et la structure modulaire à celles du modèle MA.

Cependant, les deux types de structure de réseau peuvent être observés dans les réseaux écologiques (Flores et al. 2011 ; Beckett & Williams 2013). Un réseau peut avoir une structure modulaire à une grande échelle spatiale ou taxonomique et une structure emboîtée à l'intérieur de ses modules. Cela signifie que, lorsqu'ils sont utilisés de manière complémentaire, les deux modèles, GFG et MA, offrent une description de la co-évolution de deux populations
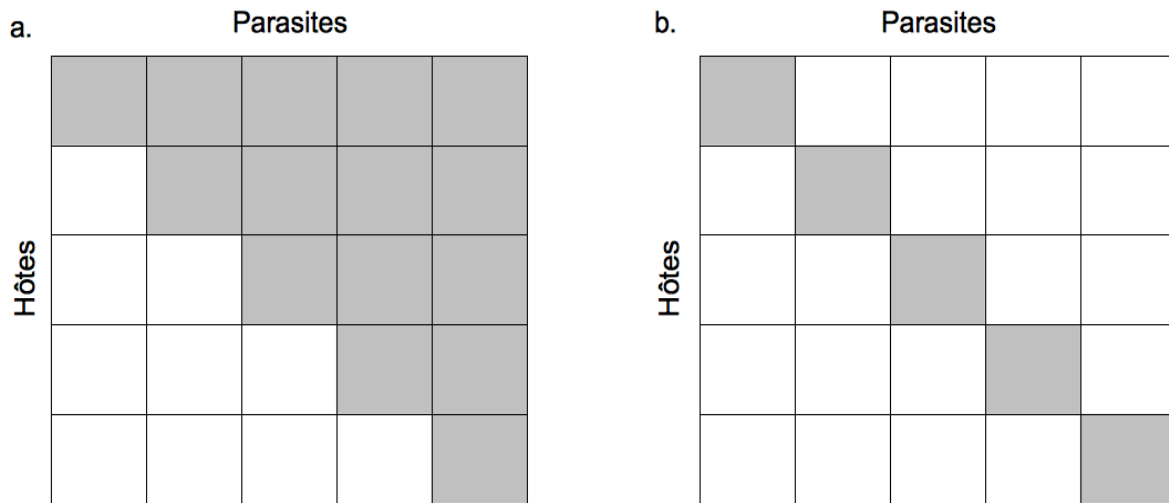
Figure 1.3 Représentation schématique des structures génétiques hôte-parasite obtenu via les modèles (a) Gene-For-Gene et (b) Matching-Allele. Les boîtes grises indiquent que le parasite est capable d'infecter l'hôte (remplacé par 1 dans les matrices d'adjacence) (a) Le parasite le plus généraliste aura la capacité d'infecter un large nombre d'hôtes, comparé au parasite le plus spécialiste, cette structure peut être comparée à cette des réseaux emboîtés. (b) Pour qu'un parasite puisse infecter un hôte, le génotype de celui-ci soit parfait compatible à celui de l'hôte. Cette structure est associée à une structure modulaire.

antagonistes qui semble plus précise que leurs prédictions prises séparément. Par exemple, un bactériophage très général sera toujours capable d'infecter la bactérie au phénotype ancestral, mais il ne sera pas aussi efficace que sur la bactérie au phénotype *coévolué* (Buckling & Rainey 2002).

1.4 Bactéries-Phages, un système bien étudié

1.4.1 Bactéries et Phages

Darwin pensait qu'il était impossible de voir l'évolution et la coévolution des populations, heureusement pour nous il n'avait pas toujours raison. La coévolution a été étudiée pour plusieurs types d'interactions (prédation, mutualisme, parasitisme), en utilisant différents types d'organismes. Les bactéries, par exemple, ont fait l'objet de nombreuses études. Elles sont

l'organisme vivant le plus abondant sur Terre et sont absolument partout. De plus, les bactéries sont cruciales pour le fonctionnement des écosystèmes. Elles contribuent en effet fortement aux cycles biogéochimiques, comme celui du nitrate ou du carbone, mais participent aussi activement à la décontamination de l'eau, à la décomposition de la matière organique, voire à la décomposition du plastique. Et grâce à leur courte durée de vie et à leur capacité à se développer facilement en laboratoire, les bactéries sont idéales pour observer et suivre la coévolution.

Parmi les organismes antagonistes aux bactéries, les parasites sont important de par l'influence que ceux-ci vont avoir sur la dynamique évolutive et démographique des bactéries. L'un d'entre eux, et qui nous intéresse ici sont les virus (voir *Glossaire*). Les virus sont omniprésents, avec environ 10^{31} d'entités estimées sur Terre (voir Breitbart & Rohwer 2005). La grande majorité d'entre eux infectent les bactéries et sont appelés bactériophages ou phages. Les phages jouent un rôle non négligeable dans le fonctionnement des écosystèmes puisqu'ils régulent les populations bactériennes. Cette régulation peut avoir pour conséquence l'augmentation la diversité des communautés bactériennes (Koskella & Brockhurst 2014), ce qui est un bon moyen d'assurer le bon fonctionnement des processus de l'écosystème (Suttle 2005). Depuis leur découverte au début du vingtième siècle par Twort et D'Hérelle séparément, de nombreuses études ont tenté de les définir, de mettre en évidence leur rôle dans l'écosystème et de comprendre la manière dont ils coévoluent avec les bactéries (Breitbart & Rohwer 2005).

De par leur facilité de culture en laboratoire, les études expérimentale portant sur les bactéries et les phages permettent de voir comment l'évolution des interactions entre les espèces évolue en fonction des conditions biotiques (*e.g.* dans des communautés de bactéries, avec plusieurs phages) et abiotiques (*e.g.* effet de la température, de la quantité de ressources, etc.). Cela représente un véritable terrain de jeu pour l'étude de la coévolution, dans l'optique de mieux comprendre son fonctionnement et les processus impliqués puisque le milieu contrôlé permet de choisir les variables d'intérêt. Ce modele biologique est aussi prompte au dévelopement de

l'utilisation de théorie des graphes dans l'étude de l'évolution, via l'utilisation de la structure des réseaux d'interactions. Malgré cette facilité expérimentale, les modèles théoriques portant sur le système bactérie-phage sont encore peu nombreux, reflétant de fait un manque de connaissances théoriques sur ce système.

1.4.2 Coévolution bactéries-phages

Parce qu'ils sont impliqués dans de nombreux processus, et pas seulement dans le domaine écologique (Scanlan et al. 2016), l'évolution et la coévolution des bactéries et des phages sont de plus en plus étudiées depuis 2000. En écologie, Buckling & Rainey (2002) a démocratisé un protocole efficace pour étudier la coévolution entre les deux. Depuis, plusieurs systèmes d'interactions bactéries-phages, en particulier *Pseudomonas fluorescens SBW25* et son phage lytique $SBW25 - \phi2$, ont été largement utilisés pour décrire et comprendre la coévolution bactéries-phages.

Contrairement à ce que l'on pourrait penser, les dynamiques de coévolution ne sont pas figées dans le temps. Le célèbre couple *P. fluorescens* - $\phi2$, par exemple, suit une ARD aux premiers stades de sa coévolution, puis passe à une FSD au fil du temps (Hall et al. 2011a). Cependant, la coévolution n'est pas conduite par la même dynamique pour chaque interaction phage-bactérie. Ces dynamiques sont en effet associées au récepteur protéique utilisé par le phage pour infecter (Betts et al. 2014), avec un jeu de compatibilité serrure/clef, induisant des dynamiques de coévolution différentes selon le phage.

La phase d'ARD de la coévolution induit l'augmentation de la résistance bactérienne et de l'infectivité des phages au fil du temps. Cette généralisation des phages conduit à des phages localement adaptés sur des bactéries sympatriques (*i.e.* de la même population) (Vos et al. 2009) mais cette relation reste peu claire (Scanlan et al. 2016). D'autre part, la coévolution semble induire une adaptation locale des bactéries aux phages sympatriques (Buckling & Rainey 2002) au détriment des allopatriques (*i.e.* de populations différentes). L'augmentation de

la résistance bactérienne aux phages sympatriques n'est pas sans coût. Il existe en effet un compromis entre résistance et fitness bactérien, causé par des mutations délétères (Buckling et al. 2006 ; Middelboe et al. 2009 ; Scanlan et al. 2015).

Toutes ces études ont contribué à l'avancement des connaissances en matière de coévolution. Ces représentations simples des interactions bactéries-phages manquent cependant une part importante d'adaptation locale et de facteurs coévolutifs, et en particulier les conditions environnementales abiotiques.

1.5   Les réseaux en contexte évolutif

1.5.1   La structure des réseaux et les dynamiques évolutives

Au-delà de l'étude de la coévolution, les réseaux peuvent être utilisés dans une multitude domaines, et ce à différentes échelles d'organisation, des individus en passant par les espèces jusqu'aux méta-communautés.

En évolution, deux niveaux sont généralement opposés : la micro- et la macro-évolution. Historiquement, les outils utilisés pour étudier l'évolution se basaient sur l'utilisation de méthodes phylogénétique. Cependant, ces méthodes nécessitent d'avoir accès à des données, soit par la récolte de fossiles, soit par des méthodes génomiques, à partir desquelles des données phylogénétiques peuvent être mesurées et/ou les arbres phylogénétiques reconstruits. Pour les niveaux supérieurs, les dynamiques démographiques sont également des indicateurs de l'évolution, avec par exemple les taux d'origination ou d'extinction, indiquant l'existence de radiation évolutive à un temps donné. Ces types d'indicateurs ont un point commun, ils sont influencés et influencent les interactions écologiques. Par exemple, au travers du conservatisme phylogénétique certaines espèces vont se retrouver à interagir avec d'autres du fait de leur histoire phylogénétique commune. Pour les hôtes-parasites, cela peut se traduire par le fait que des parasites phylogénétiquements proches risquent de partager des hôtes.

Les variations démographiques peuvent également induire une probabilité plus ou moins forte qu'une interaction se produise, dans environnement spatial et/ou temporel. De ce fait, les réseaux d'interactions d'avèrent être un outil adapté pour suivre et comprendre l'évolution, peu importe l'échelle. De plus, les réseaux évoluent eux-mêmes dans le temps et dans l'espace, ce qui peut justement apporter de nouvelles connaissances dans les mécanismes sous-jacents à l'évolution.

1.5.2   Les modèles mathématiques comme outil

Dans l'étude des réseaux d'interactions, l'utilisation de modèles théoriques se révèle être un atout de taille, puisqu'ils permettent tantôt de reconstruire, tantôt de prédire, tantôt d'expliquer ce que l'on observe dans l'environnement (Annexe A). Les données concernant les interactions n'étant pas encore des plus accessibles, ces modèles théoriques ont donc été d'une aide inestimable pour comprendre les processus impliquer dans la structure des interactions tel qu'on les observe, mais également pour comprendre leur lien avec les processus évolutifs. La diversité de modèles disponibles et de ceux à écrire en termes de focale (individus, communauté, etc.) ou encore de dynamique apporte une amplitude de mouvement dans l'étude des réseaux d'interaction dont la limite se trouve être notre imagination. L'enjeu du type de modèle en fonction de l'échelle d'étude est une chose qui a pu ressortir dans la littérature. En effet, pour des études au niveau micro, et notamment pour l'étude de la coévolution, il y aurait une tendance à utiliser davantage des modèles individus centrés. La plupart de ces modèles ne prennent en compte que deux espèces ou populations. Il y a, en effet, encore très peu de modèle individus centrés s'attelant à la tâche d'étudier l'évolution avec trois ou plus populations. À plus large échelle, pour l'étude des communautés ou plus large, les modèles sont le plus souvent basés sur des espèces que des individus (Hall et al. 2020).

1.5.3   Les réseaux pour combler un manque

L'approche par réseaux, bien qu'étant une discipline relativement récente, peut donc apparaître comme un outil idéal, notamment pour les études en évolution. En effet, les mesures utilisées par les réseaux, comme la définition de leur structure, permettent d'avoir un outil qui s'adapte à tous les niveaux : réseaux de protéines, de population, de métacommunautés, etc. Cependant, à l'heure actuelle, peu d'études se focalisent sur son utilisation en évolution et aucun consensus n'est établi concernant son utilisation exacte, la méthodologie ou l'analyse de ces réseaux dans un contexte évolutif. La validation de cette approche en évolution permettrait par exemple de développer une théorie globale sur l'évolution des espèces composant les communautés, qui est aujourd'hui encore manquante. Pour l'assemblage des communautés en particulier plusieurs théories ont été avancées, lesquelles s'apparentent plus à différents morceaux d'un même puzzle qu'à des théories complètement distinctes. Et c'est justement ici que la combinaison d'études à différents niveaux d'organisation pourrait être intéressante, puisque bien souvent ceux-ci sont étudiés de façon séparée, alors la compréhension du fonctionnement de chaque niveau est nécessaire pour avoir une idée générale claire. L'approche par réseaux offre donc la possibilité d'avoir des outils et mesures communes pour relier ces différents niveaux.

1.6   Liens entre dynamiques éco-évolutives et structure des réseaux

Le lien entre les réseaux et l'évolution est étudié depuis plusieurs décennies maintenant, mais ce n'est pas forcément le champ le plus développé en écologie-évolution. Historiquement, la majorité des interactions et réseaux d'interaction étudiés étaient des réseaux trophiques, mais ce pan la recherche, utilisant les réseaux, s'est progressivement ouvert aux autres types d'interactions. L'étude des réseaux a permis d'avoir une meilleure compréhension des dynamiques en place dans les écosystèmes. Par exemple, Fontaine et al. (2011a) ont mis en évidence que dans les communautés écologiques, les interactions et notamment leur structure, étaient des moteurs majeurs des processus évolutifs.

En ce qui concerne la structure spécifique des réseaux d'interaction, il a été mis en évidence que différentes structures étaient associées à des effets différents. Par exemple, l'emboîtement a un impact positif sur la stabilité des réseaux mutualistes, mais aussi sur leur résistance face aux perturbations. À l'inverse, l'emboîtement vient déstabiliser les réseaux antagonistes (Thébault & Fontaine 2010a). Pour ce qui est de la structure modulaire, elle aura un effet opposé dans les réseaux mutualistes et antagonistes où elle aura tendance à déstabiliser pour les premiers et stabiliser pour les seconds (Thébault & Fontaine 2010a). Cela nous montre donc que la structure des réseaux peut avoir un rôle majeur sur la façon dont les communautés sont structurées et la façon dont elles évoluent. Au niveaux des individus, donc à une échelle plus fine, Loeuille & Loreau (2005) ont montré que la modularité variait en fonction de la généralité des espèces présentes.

Si les réseaux influencent l'évolution, l'inverse est également vrai. En effet, la structure des réseaux trophiques par exemple est en partie contrainte par la phylogénie des espèces (Cattin et al. 2004a). Cela nous amène au fait que le conservatisme phylogénétique influence sur la structure des réseaux. Ce phénomène se retrouve dans les réseaux antagonistes comme mutualistes, et a permis de faire émerger la présence d'un signal phylogénétique dans les réseaux. Les interactions seraient donc déterminées par les processus écologiques (Segar et al. 2020), lesquels seraient à leur tour influencés par les interactions (Bascompte 2007). Les traits sont la pierre angulaire de la relation entre les réseaux et l'évolution. Ces traits pourraient être le lien manquant entre les études micro et macro-évolutives puisque ceux-ci sont le résultat de la macroévolution, mais ils déterminent les interactions entre les espèces au niveau micro. La sélection qui peut s'exercer sur ces traits contraint donc la structure des réseaux et donnerait donc une sorte de convergence des topologies de ces réseaux aux différentes échelles (Loeuille & Loreau 2005).

Dans le contexte évolutif, l'utilisation de modèles s'est peu à peu étendue, palliant au fait que les données sur les interactions soient parfois difficiles à récolter. Cela a joué un

rôle non négligeable dans le développement des connaissances que l'on a aujourd'hui sur le sujet. En effet, les modèles mathématiques ont la puissance de pouvoir incorporer des dimensions évolutives complexes pour essayer d'en comprendre les subtilités (Proulx et al. 2005a). Cependant, n'importe quel modèle ne convient pas à n'importe quel objet d'étude. Par exemple, les différents types de réseaux (chaine trophique, parasitisme ou mutualisme) nécessitent des approches computationnelles différentes (Eklöf et al. 2013a), tout comme l'échelle d'observation utilise différents modèles comme nous l'avons vu plus haut.

Aujourd'hui encore, il n'y a malheureusement pas de cadre général qui guide l'étude des liens entre les réseaux et les processus évolutifs.

1.7   Problématique et Objectifs généraux

La compréhension de l'évolution des écosystèmes, que ce soit au niveau de population ou de communautés, reste un domaine qui n'a pas encore livré tous ses secrets. Comme nous avons pu le voir plus haut, l'approche par réseaux est un outil puissant pour notre compréhension de l'écologie et de l'évolution. À travers cette thèse, j'essaye d'atteindre plusieurs objectifs qui sont (i) d'apporter un pont de plus entre l'approche par réseaux et l'évolution, (ii) de renforcer les liens entre la théorie et les expérimentations en matière d'évolution et plus spécifiquement de coévolution entre les bactéries et les phages, et enfin (iii) de tester l'approche par réseaux d'interactions à différentes échelles et pour différents focus en évolution. Pour cette raison, cette thèse est composée des chapitres suivants.

Le deuxième chapitre de cette thèse correspond à une introduction à l'utilisation des réseaux en écologie. Ce chapitre d'encyclopédie présente les bases de la théorie des graphes appliquées à l'écologie. Il met également de l'avant les contributions de l'approche des réseaux dans la compréhension des propriétés et du fonctionnement des écosystèmes et les potentiels développements de cette approche.

Le troisième chapitre présente une large revue de littérature sur les méthodes principalement

utilisées dans l'analyse par réseau. Elle recense les méthodes les plus robustes ainsi que leurs limitations.

Le quatrième chapitre quant à lui relate les essais d'expérimentation de coévolution entre les bactéries et les phages d'un écosystème particulier, les Sarracénies pourpre (*Sarracenia purpurea*). L'objectif premier de cette thèse était de mettre en place une expérimentation en laboratoire qui nous permettrait d'étudier l'influence de la température sur la coévolution entre les bactéries et les phages, et qui pourrait par la même occasion être reproduite en modèle mathématique. Ce chapitre illustre les démarches entreprises ainsi que les défis rencontrés, qui ont malheureusement mené à l'abandon de cette partie expérimentale.

Le modèle construit et utilisé dans le cinquième chapitre est basé sur les expérimentations en laboratoire entre bactéries et phages. Ce modèle permet de suivre l'évolution des traits des individus, bactéries autant que phages, et de recréer les réseaux d'interactions au cours du temps. Grâce à cela, il est possible d'analyser la structure des réseaux au cours du temps et identifier ce qui est susceptible de la faire changer. Ce modèle sert également à comprendre l'utilisation que nous pourrions avoir des réseaux dans le cadre de l'étude de la coévolution entre phages et bactéries en laboratoire.

Finalement, le sixième chapitre, lui aussi basé sur l'utilisation des réseaux dans un contexte évolutif, se concentre cette fois-ci sur une échelle plus large, celle des communautés. Au travers d'un modèle mathématique, nous suivons l'établissement et l'évolution de communautés en fonction du type d'interaction qui prennent place entre les individus de ces communautés. Le but ici est d'identifier l'impact du type d'interaction sur l'évolution de ces communautés, le tout en utilisant la structure des réseaux comme indicateur, mais également la structure phylogénétique de ces communautés.

Chapitre 2

Complex Ecological Networks

Complex Ecological Networks

Mathilde Besson[1,2], Eva Delmas[1,2], Dominique Gravel[2,3] & Timothée Poisot[1,2]

[1] Département de Sciences Biologiques, Université de Montréal, QC H3C 3J7, Canada; [2] Centre de la Science de la Biodiversité du Québec, Montréal, QC H3C 3J7, Canada; [3] Département de biologie, Université de Sherbrooke, QC J1K 2R1, Canada

Contribution statement:

Summary

Abstract

Ecological networks provide a useful abstraction of ecological systems, representing them as graphs, composed of nodes (species) and edges (interactions). This formalism allows to use a whole set of measures, extended from graph theory, to study ecological systems. In this chapter, we review some of the most prominent findings and areas of research from the last decade. We start by reviewing how it was used to uncover invariance in the organization of ecological systems. Then we show the importance of structure when studying systems dynamics and how this coupled approach sheds new light on emerging properties of ecological systems. Through this chapter we want to highlight the important contribution of networks in clarifying ecosystem properties and functioning, but also the potential to develop new approaches, for example to compare ecosystems and to relate species traits to community structure.

Keywords: Assembly, Coexistence, Communities, Graph theory, Network structure, Community dynamics, Species interactions, Stability

2.1 Glossary

Adjacency matrix: Matrix representing species interactions. If two species and interact, the intersection of the matrix at will be 1, and 0 if not.

Assembly rules: Ecological processes leading to a specific species' composition of a community, *e.g.* competition, predator-prey interactions, arrival history, etc.

Degree: The degree of a node is its number of links (*e.g.* interactions per species). At higher level, the degree distribution represents the cumulative distribution of links per node within the network or a subnet of the network.

Ecological interactions: Every type of contact between two species that alters the fitness of one or both species. Interactions can be directed or undirected, weighted or unweighted. They usually fall into one on these 5 main classes: competition, predation, parasitism, mutualism and

commensalism.

Ecosystem functioning: Biotic and abiotic processes that sustain ecosystems, including flues of energy and nutrients between the components of ecological systems and the resulting stocks, *e.g.* biogeochemical cycles.

Graph theory: Mathematical framework used to represent the relationship between the objects of a network.

Network structure: General shape of a network emerging from the organization of the interactions between its components. It is commonly described in ecology using connectance, link distribution, topological indices (such as nestedness, modularity, centrality), etc.

Nodes/Links, Vertices/Edges: Following graph theory, species are represented as nodes (or vertices), and interactions between them are represented by links (or edges).

Phylogenetic signal: Tendency of phylogenetically close species to have similar traits (and as a consequence, similar interactions).

Unipartite / Bipartite network: The graphical representation of the entire adjacency matrix offers an unipartite network representation (see fig. 2.1), where the hierarchy between nodes and their position into the network is not always visible. On contrary, a bipartite or k-partite network is a hierarchical representation of the network (fig. 2.2), where nodes are separated depending on their position or function into the network (*e.g.* pollinator-plant as bipartite network).

2.2   Introduction

Interactions between the components of any ecological systems are organized non-randomly. The species that form a community for example do not interact at random. The resulting organization of interactions between species drives some properties of the community such as stability, productivity, and the ability to resist extinctions, all of which eventually feedback on the organization of the system. The constant interplay between the organization of interactions

and system dynamics constrains its structure. Studying the structure of ecological systems provides insights on the fundamental rules and processes that govern ecosystem formation, maintenance and functioning.

The organization of interactions in a community is best represented as a network. *Graph theory* is a field of mathematics developed to analyze the structure of such systems. Every community can be abstracted by a *graph*, which is a representation of the system components and their arrangement (fig. 2.1 a). These components are called *nodes* and are linked together by *edges*. In an ecological system, nodes can be individuals, populations, communities or landscape patches and edges can represent trophic interactions, energetic flues and more generally every kind of interactions. Both nodes and edges can carry additional information such as weight (*e.g.* species abundance, intensity of the gene flow between two populations, etc.), location in space and time, and labels (*e.g.* species identity). Specific information can be attached to edges, modifying the characteristics of the graph, *e.g.* the environmental dependence of an interaction. Graphs can be *directed* (*i.e.* interaction goes from A to B) or *undirected*, *weighted* (*i.e.* different strength of interaction among the network) or *unweighted* (fig. 2.1 et 2.2). This information is summarized in the *adjacency matrix*, typically named (fig. 2.1 b). The adjacency matrix can be used to answer various ecological questions. Using it directly allows to follow direct interactions and the network structure, and using the inverse of can be useful to obtain indirect interactions, and even more (Montoya et al. 2009). In this chapter, for simplicity, we will focus mostly on *Species Interaction Networks* (SIN). Ecological systems such as landscape, genetic or nutrient networks are not represented here, but they can be studying using the same framework as defined further. Describing and understanding the structure of SIN is an active, and growing, field of ecological research. We provide here an overview of some of the most prominent findings and areas of research from the last decade. Starting from a discussion of some invariant properties of the structure of species interaction networks, we will then discuss how this structure affects community dynamics and properties.

We will follow by a discussion of the ways ecological networks can be studied under familiar concepts from ecological theory, and finally how this approach scales up to larger temporal and spatial scales.

2.3   Invariants in ecological networks

One striking particularity of ecological networks is their consistency: even though they depict interactions between different organisms across all sorts of ecosystems, they all tend to look the same (Jordano et al. 2003). Remarkably, even when interactions among species themselves vary, the overall network structure tends to remain unchanged (Kemp et al. 2017). Most ecological networks have a very specific and similar degree distribution (Williams 2011) (fig. 2.1 d), whereby most species have a small number of interactions, and a small proportions of species have a large number of interactions. In food webs, which represent interactions between prey and their predators, there is a well-described relationship between the number of species and the number of interactions. The number of interactions (L) increases proportionally to the number of species (S) raised to some exponent, or $L \propto S^k$. Martinez (1992) suggested that this exponent is approximately equal to 2, *i.e.* the number of interactions is proportional to the squared number of species. Brose et al. (2004) showed that this relationship holds even across space; it is possible to estimate how many interactions a species will establish across its entire range. In other instances, networks may differ on some aspects of their structure, despite obeying to a shared underlying principle. For example, Fortuna et al. (2010) showed that in networks with a low connectance (fig. 2.1 c), nestedness (the degree to which the diet of specialists and generalists overlaps – fig. 2.2) and modularity (the tendency of species to form densely aggregated clusters – fig. 2.2) are positively correlated. In networks with higher connectance, this becomes the opposite: networks with a large number of interactions are either nested (and not modular) or modular (and not nested). In the recent years, it emerged that many aspects of network structure covary with connectance (Poisot & Gravel 2014;

Chagnon 2015), suggesting that simply knowing how many species there are, and how many interactions they establish, is already very informative about the network structure.

Another remarkable generality of network structure is the distribution of particular inter-connection between three-species subsets. Milo et al. (2002) found that networks (not just ecological but other types of networks such as neuronal or electronical networks as well) can be characterized by the over or under representation of some of these three-species subsets, which they called motifs (fig. 2.1 e). Motifs can be more broadly defined as specific arrangements of interconnection between three (or more) nodes. The frequency at which they occur in a network can be computed and compared to randomized networks in order to reveal significant aspects of the structure. Three-species motifs represent the simplest building blocks of networks, and more importantly typical interaction modules found in communities. As such, they offer the possibility to integrate and test theories developed with simple modules in larger, more realistic networks (*e.g.* omnivory, McCann et al. (1998), Holt (1997)). Food webs, for example, are characterized by an over representation of linear food chains and omnivory and an under representation of apparent and exploitative competition (fig. 2.1 a, e) (Bascompte & Melián 2005; Camacho et al. 2007). Stouffer & Bascompte (2010) found that realistic motif distribution promotes stability in food webs, with over-represented motifs being more stable in isolation and correlated with higher stability in large realistic communities, and conversely. Motifs can also be used to characterize species role in networks. From the 13 different three-species motifs emerge 30 unique positions for species to occupy in these motifs, representing how the species is embedded in its community. The different positions a species will occupy, and the frequency with which it will occupy these different positions in networks are called species motif role (Stouffer et al. 2012). These roles have been shown to be evolutionary conserved in food webs (Stouffer et al. 2012) and to have less variability in time than expected in host-parasitoids bipartite networks (Baker et al. 2015).

Another invariant network property relates to evolutionary history. Phylogeny is a key
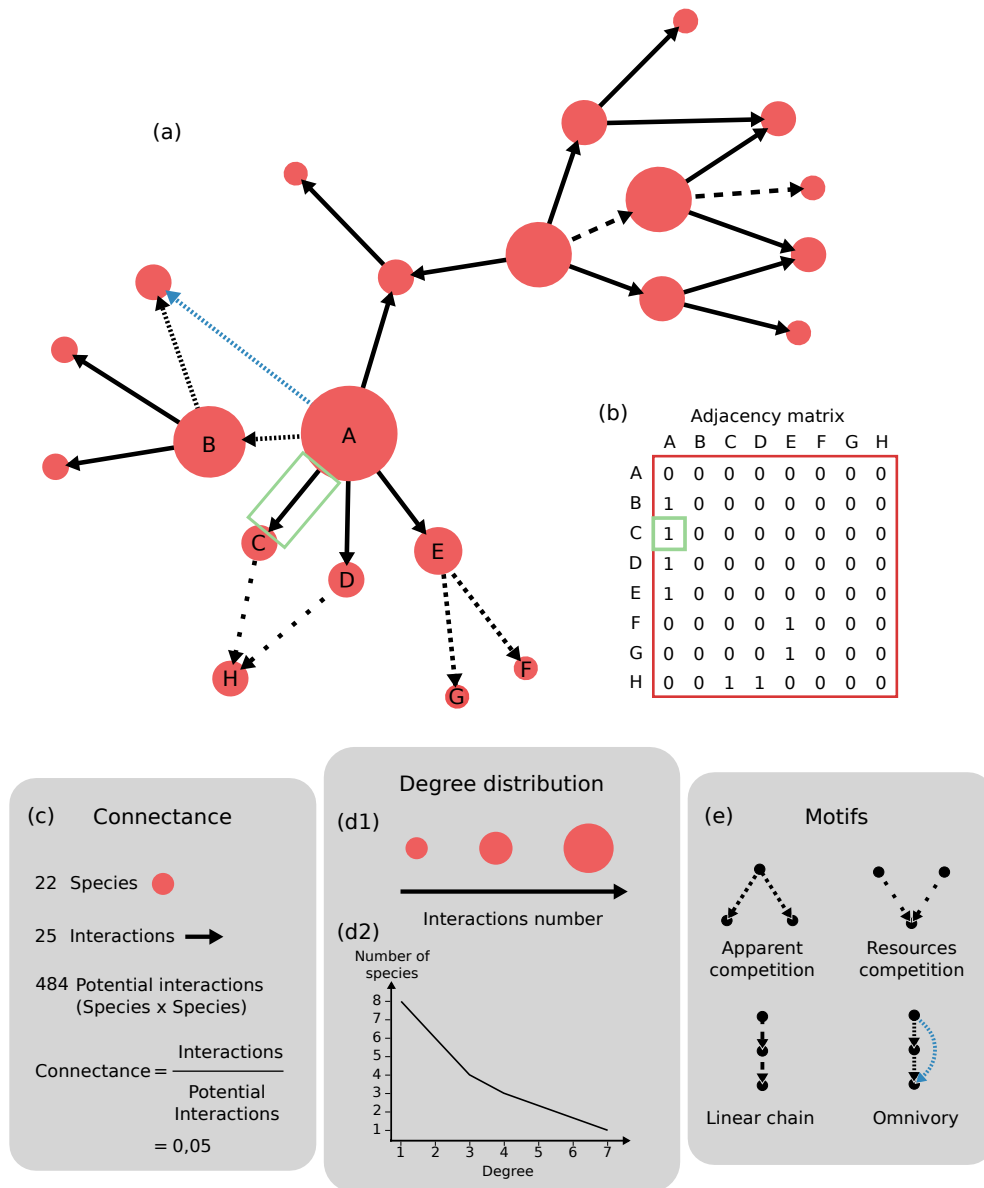
Figure 2.1 Graphical representation of an ecological network (A), where species are represented by circles and their directed interactions by arrows. The representation is formalized in the adjacency matrix (B). In an unipartite representation as this one, each species is represented both as a column and a row. 1 indicates an interaction between two species (e.g., the green square in (B)), and 0 indicates the absence of interaction. This matrix facilitates computation of characteristics such as the connectance (C) and the degree distribution (D). (C) Represents the level of connection into the network and is calculated as showed in the figure. (D) Represents the distribution of interaction per species. The circles size is relative to the amount of interactions a species have (D1). This distribution is nonrandom and generally follows a power-law distribution (D2). The network can be split into subnets composed of 3 species, called motif (E). Among the 13 different possible motifs, we only represented the most commonly found in natural communities.
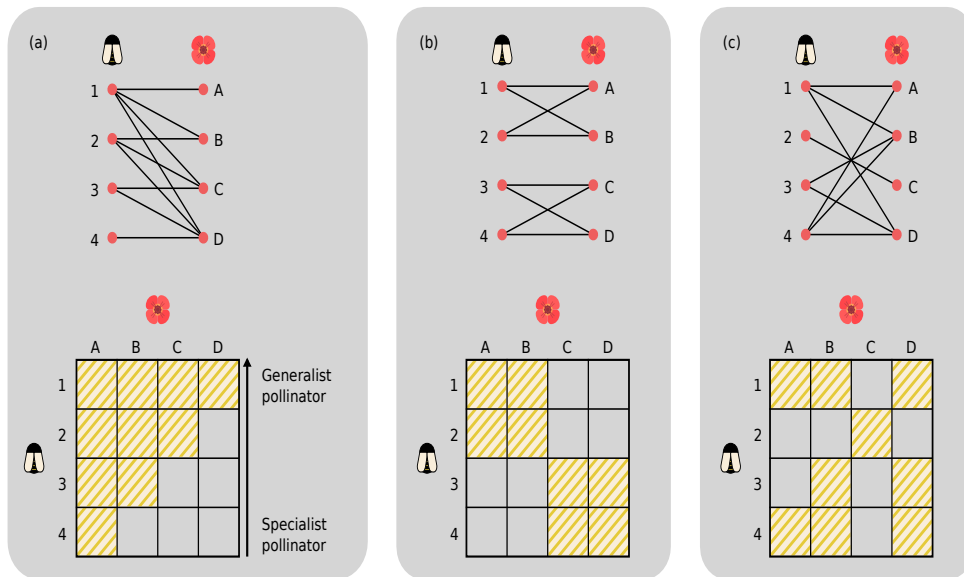
Figure 2.2 Network topology, example of a fictional plant–pollinator network. (A) shows a perfectly nested network, where specialists pollinators are visiting plants embedded into the diet of more generalist pollinators. (B) Shows a perfectly modular network, where subgroups of species interact more strongly with each over than with the rest of the network. (C) Shows a random network. Two representations are possibles. Top: Bipartite representation using nodes and edges; Bottom: Ordered interaction matrix. Here, we used striped yellow squares instead of 1 for presence of interaction and empty squares in absence of interaction.

determinant of ecological network structure, being related to species position and interactions into the community. Phylogenetically close species tend to inherit traits from their common ancestors (*e.g.* body size, habitat, defensive strategy, metabolic type, phenology), increasing their propensity to interact with the same group of species or with similar species, a phenomenon called phylogenetic signal. This conservatism of interactions has been found to hold across different types of interactions such as antagonistic or mutualistic interactions (Fontaine & Thébault 2015). However, depending in the species role (*e.g.* host or parasite, pollinator or plant) the link organization will be different, leading to an asymmetrical structure for pairwise interactions. For instance, closely related hosts tend to share parasites, while closely related parasites, because of competition for resources, tend to have different hosts (Krasnov et al. 2012). The conservatism of interactions is consequently unequal all over the network. Following the logic that closely related species interact with the same group of species, Rezende et al. (2009) showed that phylogenetic structure of ecological networks explains almost entirely the formation and composition of modules and the connections between them. The species connecting modules together are indeed usually phylogenetically close. Cattin et al. (2004a) also found, using a niche-hierarchic model, that diet is constrained by the phylogenetic origin of consumers. The nested structure of trophic networks is then influenced by the phylogenetic signal of interacting species and their traits compatibility. In contrast, the nested structure of mutualistic networks would be a consequence of trait complementary between species (Rezende et al. 2007). For now, mechanisms underlying the nestedness-phylogeny relationship remain to be further investigated. Moreover, because of species plasticity, phylogeny alone does not fully explain the structure and evolution of ecological networks.

2.4 From structure to properties

The relationship between ecological network structure and stability is a long-lasting object of research in community ecology. MacArthur (1955) and Elton (1958) first proposed that diverse

communities should have a more stable dynamic than simple ones because disturbances are more easily spread through highly connected nodes. May (1972) countered this hypothesis using a mathematical model based on random ecological networks and proposed there should be a limit to ecosystem complexity. This counter-intuitive proposition sparkled live debates still lasting today (McCann 2000; see Allesina & Tang 2015). Two different approaches to the problem followed: one focused on dynamical stability and the other on the resistance of communities to species lost. Despite their dissimilarities, these approaches are not totally independent (Donohue et al. 2013) and revealed that species diversity has no direct influence on community stability. However, the structure of ecological network such as the distribution of interaction strength and network topology seems to play a crucial role (Yodzis 1981).

As mentioned above, the degree distribution of ecological networks often follows a power-law distribution (Montoya & Solé 2002), indicating that few species are highly connected to the rest of the community and a large number of species are weakly connected to others. This organization combined with the myriad of weak interactions found across ecological networks buffers species variations and stabilizes the dynamics of the entire community (Bascompte et al. 2005; Jacquet et al. 2016). Other aspects of community structure, such as the predator-prey body-mass ratio (Emmerson & Raffaelli 2004; Brose et al. 2006a) and network architecture (Montoya et al. 2006; Thébault & Fontaine 2010b), determine the distribution and strength of interactions and together drive the stability of ecological networks (Jacquet et al. 2016).

Perturbations in ecological communities such as landscape fragmentation, habitat loss, or species invasion, are the primary drivers of species loss. Extinctions may happen directly, for instance if a particular habitat is eliminated, or indirectly following a first species loss (a phenomenon referred as secondary extinction or cascades). Such extinctions are used to measure the robustness of ecological communities. Simulation experiments revealed that the likelihood of secondary extinctions increases with community size ((Lundberg et al. 2008), decreases with network connectance (Dunne et al. 2002b) and primarily affects the most

isolated species in the network. The loss of a highly connected species, also called a hub, induces a higher rate of secondary extinctions than the loss of a random and weakly connected species (Solé & Montoya 2001). Similarly, species responsible for important energy-flow in the network (carbon, nitrogen or biomass) can trigger secondary extinctions (Allesina & Bodini 2004).

The network architecture also affects the community response to perturbations. In agreement with MacArthur's intuition, it was found that species with low degree also more strongly propagate perturbations following permanent changes in the environment because of their tight connections (Montoya et al. 2009). Alternatively, the most connected species diffuse such perturbations through the network and even though they affect a higher number of species, their average effect on other ones is much smaller. Overall network properties also affect the response to perturbation. Thanks to their structural properties (high nestedness and connectance, Jordano et al. 2003), mutualistic networks persist longer than randomly structured networks (Memmott et al. 2004; Fortuna & Bascompte 2006). On the other hand, presence of modules in the community structure limits propagation of perturbations across the rest of the network and, as such, secondary extinctions (Stouffer & Bascompte 2010).

Elucidating the consequences of biodiversity lost for ecosystem functioning is also an important field where the network approach has been useful. The hypothesis that an increase in species diversity results in an increased productivity dates back to (Darwin et al. 1859) and a formal theory for what is now called the biodiversity-ecosystem functioning (BEF) relationship was proposed in the mid 90s. In a trophic group (*i.e.* a group of species that all belong to the same trophic level, e.g. producers or herbivores), increasing diversity improves resource use efficiency and translates into larger productivity (Loreau 2010) (*e.g.* nutrients for producers, or producers for herbivores). Yet, when the trophic group under focus is coupled to other(s), the action of diversity on functioning is more variable (Duffy et al. 2007). This makes the BEF relationship unpredictable in real-world communities (Harvey et al. 2013), composed of several

trophic groups that are virtually never differentiable – as intraguild predation and omnivory blur the frontier between levels. The multiplicity of the factors influencing the BEF relationship calls for a more general framework that allows the integration of the theories developed for trophic groups and for simple modules or sub-systems (Gravel et al. 2016a). By mapping transfer of biomass and energy and/or constraints on organism through the different compartments that compose a natural community, ecological networks – and food webs in particular – offer the possibility to perform this integration. Analyses performed on simulated food-webs with fixed species richness have shown that interactions, and more specifically their structure, have a significant influence on productivity (Thébault & Loreau 2003; Thébault et al. 2007; Poisot et al. 2013). The structure of interactions is indeed a reflection of community properties, essential to ecosystem functioning. It seems then essential to integrate it in BEF studies.

2.5   Mechanisms underlying pairwise interactions

Ecological interactions between species should be viewed as the result of low level processes involving pairs of individuals. A pollinator is able to effectively reach the nectar in a plant because their respective traits match, they have compatible phenologies, and they occur in the same environment. A virus can infect its host because it is able to attach to the cell surface, effectively penetrate it, and hijack the cellular machinery to its benefit. Interactions that are not allowed because trait values do not match have been called "forbidden links" (Olesen et al. 2011). This prompted a search for "linkage rules" (Bartomeus 2013) in ecological networks, *i.e.* the relationships that must exist between traits of two organisms in order for an interaction between them to exist. These can be identified from existing data on traits and interactions (Bartomeus et al. 2016), and then used to generate realistic ecological networks (Crea et al. 2016). González-Varo & Traveset (2016) pointed out that interactions are happening between individuals, and as a consequence, it requires to consider not only how the traits are distributed at the individual scale, but also how different behaviors may allow organisms to overcome some

of the forbidden interactions.

Although traits are an important part of what makes interactions happen, they are only relevant insofar as the organisms are able to encounter one another. The importance of neutral dynamics (*i.e.* how abundances of different species can determine the probability that they can interact, based on how often they would get in contact by chance) is, somewhat counter-intuitively, great. Canard et al. (2012) revealed that realistic food webs can be predicted with only knowledge of abundances. In a host-parasite system, local abundances has also been identified as a key predictor of species interactions (Canard et al. 2014). More broadly, because interactions emerge from all of these ecological mechanisms, there is a need to develop a deeper understanding of their variability (Poisot et al. 2015c). Beyond the fundamental advance that this represents, this would allow to model interactions based on external information instead of documenting all of them (Morales-Castilla et al. 2015).

The realization of an interaction between individuals has, by definition, an effect on population dynamics. But it is also archetypical of complex system dynamics, where low level processes propagate up to higher level of organization and impact emerging properties of the community. If we consider for instance a population A, its dynamic is not the same when it multiplies in isolation – where it can grow exponentially if resources are unlimited (Malthus 1798) or logistically otherwise (Verhulst 1938) – or when it is embedded in a real-world community, composed of several species interacting with one another through different processes. That population can lose individuals to predation, have parasitism increase its death rate and at the same time see its establishment eased through facilitation. It then becomes necessary to account for the entire set of interactions to understand population, community and ecosystem dynamics. But the effect of interactions on dynamics is not always straightforward to infer, both in terms of directionality and intensity, as there is different types of interactions and multiple factors influencing their occurrence and strength.

Ecological networks are also spatially and temporally variable (Trøjelsgaard & Olesen 2016).

There are two drivers to this variability: changes in species composition, and changes in the way these species interact (Poisot et al. 2012). Changes in species alone are able to generate variation in network properties (Havens 1992). Spatial variation in network structure can also reflect deep-time constraints; for example, Dalsgaard et al. (2013) revealed that historical climate change trends have a signature on the nestedness and modularity of pollination networks. Even when the same species are present, interactions between them can vary. Carstensen et al. (2014) and Trøjelsgaard et al. (2015) investigated this phenomenon in mutualistic networks. Interaction turnover results from variations in partner fidelity (some species pairs are extremely closely associated), but also from variations in the local environment in which the species interact. Interestingly, networks overwhelmingly tend to conserve their structure even when interactions within them change. Díaz-Castelazo et al. (2010) surveyed a pollination network over 10 years, and found important species turnover during this period. Nevertheless, the network retained its structure because species were replaced by their functional equivalent; a generalist pollinator often succeeded to another generalist pollinator. Conversely, species tend to retain their role in different communities: Baker et al. (2015) showed that species keep occupying the same position in the network across space, regardless of the species they interact with at every location.

2.6   From the regional species pool to local structured communities

Describing the variation in ecological network structure at large spatial scales may represent an additional layer of information compared to simple species lists. As such, ecological networks are a powerful tool to shed new light on the processes underlying species distribution (Cazelles et al. 2016) and variation in some ecosystem functions (*e.g.* trophic regulation). Until recently, the prevailing idea was that at large spatial scales, the role of biotic interactions on distribution is very small compared to that of abiotic conditions, and as such is important only locally (Pearson & Dawson 2003; Boulangeat et al. 2012). Empirical observations of species-environment

relationship are used to approximate species physiological tolerance to environmental conditions and potentially predict their range under different scenarios of climate change (*e.g.* Araújo et al. 2006). While these species distribution models provide a useful approximation of their potential range shift (Pearson et al. 2002), there is mounting evidence that biotic interactions – both positive and negative – play a critical role in shaping communities not only at local scales (Boulangeat et al. 2012), but also at macro-ecological scales (Davis et al. 1998; Araújo & Luoto 2007; Heikkinen et al. 2007; Gotelli et al. 2010; Araújo et al. 2011).

It was proposed that the role of interactions in shaping species distribution could be approximated from knowledge of species co-occurrence (Araújo et al. 2011). This very active field of research has been recently pushed by the development of joint species distribution models (JSDM), which account simultaneously for the effect of the environment and co-distribution (Pollock et al. 2014). But there are limitations to this approach. For instance, it does not allow to distinguish between co-occurrence caused by biotic interactions and correlated responses to unmeasured environmental variables (Pollock et al. 2014). Conversely, the lack of association between species is not an evidence of absence of interaction (Cazelles et al. 2016). Further work is therefore needed to move from correlative species distribution models (SDM) toward more theoretically sound models. In particular, developing methods allowing to include prior information about the underlying ecological network when estimating (J)SDM could shed light on the fundamental processes underlying species distribution and thus making more accurate predictions (Cazelles et al. 2016). Additionally, Poisot et al. (2017) recently showed that biotic interactions respond to environmental conditions on their own, independently of species.

Ecological networks also offer an ideal framework to study the conditions for the maintenance of biodiversity in communities. The competitive exclusion principle states that the number of coexisting species should be equal or smaller than the number of resources. This stands in contradiction with the existence of ecological communities containing species that overlap in some extent in their resources or consumers. Phytoplanktonic communities are often considered

to illustrate this paradox (Hutchinson 1961), as they exhibit a high biodiversity while species are competing for a limited number of shared resources (*e.g.* light, nitrate). Species coexistence mechanisms (Chesson 2000) are based on species traits that either decrease fitness differences (equalizing mechanisms) and/or increase niche differentiation between species (stabilizing mechanisms).

The coexistence theory and the representation of ecological communities as networks of interactions has brought new perspective on species coexistence. Martinez et al. (2006) for instance showed that the global non-random structure of the food webs improve community persistence (*i.e.* species coexistence). The distribution of motifs in food webs (Stouffer and Bascompte 2010, see section Invariants in ecological networks) as well as species' role within motifs (Stouffer et al. 2012) are related to community persistence. In mutualistic networks for instance, the nested structure minimizes interspecific competition and increase the number of coexisting species (Bastolla et al. 2009; Sugihara & Ye 2009). Interactions structure also tend to impact species coexistence into communities, as highlighted by (Bascompte et al. 2006), the fact that one species *A* depends strongly on another species *B* as resource for food or pollination, and the other species, *B*, only weakly depends on *A*, also called asymmetry of dependences, increases coexistence of species. As an other example, using food web structure Brose et al. (2006b) showed that the allometric scaling of metabolic rates of species improve community persistence. All these types of approach, whether they are based on motifs, species' role or allometric scaling, have highlighted the importance of network structure in species coexistence.

Ecologists have also questioned the way communities are formed and the hypothetical set of rules embedding their assembly. The network approach allows to explore in details the different processes influencing ecological communities assembly. Capitán et al. (2009), for instance, characterized the sequence of species arrival in a community with an assembly graph. It allows to follow step by step every possible path in community assembly from 0 to *n* species among

several trophic levels, and to highlight underlying mechanisms. (Verdú & Valiente-Banuet 2008), for instance, found that nested community provides generalists species which facilitate the presence of other species into the network. At the same time, Olesen et al. (2008) observed that newly arriving species tend to interact more easily with already well-connected or generalist species. Such results could let us think about the Drake's controversial idea that species arrival history would be an important factor driving community assembly (Drake 1991). This proposition was supported by network analyses, such as in Campbell et al. (2011) for mutualistic networks, but still remains object of debate.

The addition of ecological networks into models of diversity dynamics fostered the development of theory of community assembly at both, fine and large spatial scales. Niche and neutral theories dominated most of community assembly research since the publication of HUBBELL (2001). A wide range of models have been used, most of them with very abstract and phenomenological representations of the niche. But only recently, with the addition of trophic constraints (Gravel et al. 2011) and other types of interactions (Cazelles et al. 2016) to MacArthur & Wilson (1967) model of island biogeography, that all types of interactions were considered in the process of community assembly. The model was first extended by assuming that predators could only colonize communities with prey already present, and go extinct with their last prey. This modification was sufficient to explain the observation of a sequential construction of food webs after the defaunation treatment of the famous experiment by Simberloff & Wilson (1969), Petchey et al. (2008b). The model was further use to illustrate a reciprocal feedback between colonization-extinction dynamics and local food web dynamics, where properties of the regional food web constrain the development of the local motif structure, and alternatively local dynamics influence the assembly process (Massol et al. 2017). This modeling approach allows a general representation of the niche in studies of assembly dynamics (Jacquet et al. 2017) and propose a unifying framework to explain the construction of local communities from a sample of the regional species pool.

2.7 Conclusion

Graph theory delivered important scientific discoveries, such as improved understanding of breakdown of electricity distribution systems or the propagation of infections in social networks. It is also a powerful tool to investigate key questions in ecology. Graph theory provides a remarkably simple way to characterize the complexity of ecological networks. Indices such as connectance, degree distribution or network topology serve as basic measurements to describe their structure. Such indices facilitate comparison between different systems and reveale commonalities and variations. Nowadays, the relatively large number of network studies leads to a myriads of ways to sample, analyze and interpret them (see Delmas et al. 2019).

Studying ecological networks have however a larger purpose than just their description and classification. Basic measurements are correlated to several environmental conditions and network analysis appears to be helpful in different ecological fields. As we seen through this chapter, it can be used to study dynamics of ecological systems and their responses to changes, according to their stability over time or the BEF relationships in the system. It also highlights the understanding of mechanisms underlying ecological properties such as community assembly, coexistence and species distribution. Network studies were a key to reveal relationships between different properties of ecological network such as trait and structure.

Chapitre 3

Analyzing ecological networks of species interactions

Analyzing ecological networks of species interactions

Eva Delmas[1,2], Mathilde Besson[1,2], Marie-Hélène Brice[1,2], Laura A. Burkle[3], Giulio V. Dalla Riva[4], Marie-Josée Fortin[5], Dominique Gravel[2,6], Paulo R. Guimarães Jr.[7], David H. Hembry[8], Erica A. Newman[9,10], Jens M. Olesen[11], Mathias M. Pires[12], Justin D. Yeakel[13,14] and Timothée Poisot[1,2]

[1] Département de Sciences Biologiques, Université de Montréal, Montréal, H2V 2J7, Canada; [2] Québec Centre for Biodiversity Sciences, McGill University, Montréal, H3A 1B1, Canada; [3] Department of Ecology, Montana State University, Bozeman, MT 59715, U.S.A.; [4] Beaty Biodiversity Research Centre, University of British Columbia, Vancouver, V6T 1Z4, Canada; [5] Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, M5S 3B2, Canada; [6] Département de Biologie, Université de Sherbrooke, Sherbrooke, J1K 2R1, Canada; [7] Departamento de Ecologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, 05508-090, Brazil; [8] Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, U.S.A.; [9] School of Natural Resources and Environment, University of Arizona, Tucson, AZ 85721, U.S.A.; [10] Pacific Wildland Fire Sciences Laboratory, USDA Forest Service, Seattle, WA 98103, U.S.A.; [11] Department of Bioscience, Aarhus University, Aarhus, 8000, Denmark; [12] Departamento de Biologia Animal, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Campinas, 13083-862, Brazil; [13] Life and Environmental Sciences, University of California Merced, Merced, CA 95343, U.S.A.; [14] Santa Fe Institute, Santa Fe, NM 87501, U.S.A

Contribution statement: ED, MB, MHB and TP wrote the manuscript outline and the first draft. ED, MB and TP contributed equally to this manuscript first draft. ED revised the manuscript after peer review. All authors besides MB, ED, MHB, GVDR and TP listed in alphabetic order. All other authors contributed to edits and discussion.

Abstract

Network approaches to ecological questions have been increasingly used, particularly in recent decades. The abstraction of ecological systems – such as communities – through networks of interactions between their components indeed provides a way to summarize this information with single objects. The methodological framework derived from graph theory also provides numerous approaches and measures to analyze these objects and can offer new perspectives on established ecological theories as well as tools to address new challenges. However, prior to using these methods to test ecological hypotheses, it is necessary that we understand, adapt, and use them in ways that both allow us to deliver their full potential and account for their limitations. Here, we attempt to increase the accessibility of network approaches by providing a review of the tools that have been developed so far, with – what we believe to be – their appropriate uses and potential limitations. This is not an exhaustive review of all methods and metrics, but rather, an overview of tools that are robust, informative, and ecologically sound. After providing a brief presentation of species interaction networks and how to build them in order to summarize ecological information of different types, we then classify methods and metrics by the types of ecological questions that they can be used to answer from global to local scales, including methods for hypothesis testing and future perspectives. Specifically, we show how the organization of species interactions in a community yields different network structures (*e.g.*, more or less dense, modular or nested), how different measures can be used to describe and quantify these emerging structures, and how to compare communities based on these differences in structures. Within networks, we illustrate metrics that can be used to describe and compare the functional and dynamic roles of species based on their position in the network and the organization of their interactions as well as associated new methods to test the significance of these results. Lastly, we describe potential fruitful avenues for new methodological developments to address novel ecological questions.

3.1 Introduction

Al-Jahiz was perhaps the first scientist to provide, as early as in the eighth century, a description of a food chain (Egerton 2002). About a thousand years later, Camerano (1880) introduced the idea that the diversity of animal forms, and therefore biological diversity itself, can only be explained when framed in the context of interrelationships among species. Seminal work by Patten (1978) and Ulanowicz (1980) suggested that the structure of networks can approximate information on theoretical constraints on community assembly, and helped generate interest in the application of network science to ecology. "Network-thinking" now permeates studies in ecology and evolution (Proulx et al. 2005b), and is one of the fastest growing ecological disciplines (Borrett et al. 2014), accounting for 5% of all published papers in 2012. Network-based approaches are gaining momentum as one of the most helpful tools for the analysis of community structure (Poisot et al. 2016b), because they offer the opportunity to investigate, within a common formal mathematical framework, questions ranging from the species level to the community level (Poisot et al. 2016b). Applying network approaches to a variety of ecological systems, for example hosts and parasites (Poulin 2010), or bacteria and phage (Weitz et al. 2013), yields new methodological and biological insights, such as the observation that networks tend to be locally nested but regionally modular (Flores et al. 2013), which suggests that different ecological and evolutionary regimes are involved at different scales. Despite this long-standing interest, the application of measures grounded in network science is still a relatively young field (in part because the computational power to perform some of these analyses was largely unavailable in the early days of the field). This comes with challenges to tackle. First, there is a pressing need for additional methodological developments, both to ensure that our quantitative analysis of networks is correct, and that it adequately captures the ecological realities that are, ultimately, of interest. Second, we need to understand better the limitations and domain of application of current methods. Yet, there is a lack of a consensus on what constitutes a "gold standard" for the representation, analysis, and interpretation of

network data on ecological interactions within the framing of specific ecological questions; *i.e.* which of the many available measures actually hold ecological meaning. All things considered, the analysis of ecological networks can be confusing to newcomers as well as researchers who are more well versed in existing methods.

Most notions in community ecology, including the definition of a community (Vellend 2010; Morin 2011), and several definitions of a niche (Holt 2009; Devictor et al. 2010), emphasize the need to study the identity of species and their interactions simultaneously (although ecological network analysis can be critiqued for ignoring species identity in many instances). Studies of ecological communities can therefore not discard or disregard interactions (McCann 2007), and using network theory allows researchers to achieve this goal. With the existence of methods that can analyze (large) collections of interactions, this approach is methodologically tractable. Graph theory provides a robust and well formalized framework to handle and interpret interactions between arbitrarily large (or small) numbers of species. Theoretical analyses of small assemblages of interacting species (*e.g.*, "community modules", Holt 1997) have generated key insights into the dynamics of properties of ecological communities. We expect there is even more to gain by using graph theory to account for structure at increasingly high orders of organization (*e.g.*, more species, larger spatial or temporal scales), because there is virtually no upper bound on the number of nodes (species) or edges (interactions) it can be applied to, and theory on large graphs can help predict the asymptotic behaviour of ecological systems. In short, although graph theory may appear as overwhelmingly complicated and unnecessarily mathematical, it allows us to express a variety of measures of the structure of networks that can be mapped onto ecologically relevant questions.

Applying measures from network science to ecological communities can open three perspectives (Poisot et al. 2016b). First, the multiplicity of measures confers additional tools to describe ecological communities. This can reveal, for example, unanticipated ways in which communities differ. Second, these measures can provide new explanatory variables to explain how ecological

communities function. The question of stability, for example, has been approached through the analysis of empirical food webs to question long-standing theoretical results (Jacquet et al. 2013). Finally, and this is a new frontier in network studies, they open the ability to predict the structure of ecological communities, through the prediction of interactions (Desjardins-Proulx et al. 2017; Stock et al. 2017). The domain of application of ecological networks is as vast as the domain of application of community ecology; but ensuring that network measures deliver their full potential of advancing our understanding of ecological systems requires that they are well understood, and well used. Because of advances in graph theory, and the availability of more efficient computational methods, the exploration of large networks is now feasible. While this may not be immediately useful to macrobe-based research, microbial ecology, through sequencing, is able to generate data sets of immense size that can be analysed with the tools we present here (Faust & Skvoretz 2002).

This review provides an assessment of the state of methodological development of network science applied to ecological communities. Taking stock of the tools available, their strengths and limitation, is a necessary first step to determine how we can best analyse data from ecological networks and improve in the future our analyses of their consequences on dynamic ecological processes (see Jordano & Bascompte 2013 for mutualistic systems; Poulin 2010 for parasites; McCann 2012 for food webs; or Dormann et al. 2017 for a recent overview). In this review, we highlight areas in which future research is needed, so as to eventually establish a comprehensive framework for how ecological networks can be analysed. The measures presented herein do not represent all the measures that are available for ecological networks; instead, they represent a core set of measures that are robust, informative, and can be reasoned upon ecologically. While this review does not present the entire framework for ecological network analysis, we are confident that it provides a solid foundation for its future development, and that the recommendations we lay out should be used by future studies. We have organized the measures by broad families of ecological questions. What is the overall structure of ecological

networks? How can we compare them? What are the roles of species within networks? How similar are species on the basis of their interactions? How can we assess the significance of measured values? What are emerging questions for which we lack a robust methodology? This order mimics the way networks are usually analysed, starting from community-level structure, and going into the species-level details.

3.2   What are species interaction networks?

Identifying interactions across ecological entities can be done in a variety of ways, ranging from literature survey and expert knowledge, direct or indirect observation in the field using gut content (Carscallen et al. 2012), stable isotopes, molecular techniques such as meta-barcoding and environmental DNA (Evans et al. 2016; O'Donnell et al. 2017), to modelling based on partial data or mechanistic models. Depending on how they were assembled, species interaction networks can represent a multitude of ecological realities. When based on field collection (Morand et al. 2002; Bartomeus 2013; Carstensen et al. 2014), they represent realized interactions, known to have happened (unreported interactions can be true or false absences, depending on sampling effort among other things). Another common method is to 'mine' the literature (*e.g.*, Havens 1992; Strong & Leroux 2014) or databases (Poisot et al. 2015b), to replace or supplement field observations. In this situation, species interaction networks describe potential interactions: knowing that two species have been observed to interact once, there is a chance that they interact when they co-occur. Another more abstract situation is when interactions are inferred from a mixture of data and models, based on combinations of abundances (Canard et al. 2014), body size (Gravel et al. 2013; Pires et al. 2015), or other traits (Crea et al. 2015; Bartomeus et al. 2016). In this situation, species interaction networks are a prediction of what they could be. In keeping with the idea of 'networks as predictions', a new analytical framework (Poisot et al. 2016a) allows working directly on probabilistic species interaction networks to apply the family of measures presented hereafter.

Interactions are compiled and resolved (and subsequently assembled in networks) for a multitude of taxonomic and organisational levels (Thompson & Townsend 2000): individuals (Araújo et al. 2008; Dupont et al. 2009, 2014; Melián et al. 2014); species (Morand et al. 2002; Krasnov et al. 2004); at heterogeneous taxonomic resolutions, including species, genera, and more diffusely defined 'functional' or 'trophic' species (Martinez et al. 1999; Baiser et al. 2012); groups of species on the basis of their spatial distribution (Baskerville et al. 2011). This is because species interaction networks are amenable to the study of all types of ecological interactions, regardless of the resolution of underlying data: mutualistic, antagonistic, competitive, and so on. Recent developments made it possible to include more than one type of interaction within a single network (Fontaine et al. 2011b; Kéfi et al. 2012), allowing greater ecological realism in representing communities, which encompass several types of interactions (*e.g.*, plants are consumed by herbivores, but also pollinated by insects). Such networks are instances of multigraphs (in which different types of interactions coexist). Another development accounts for the fact that ecological interactions may have effects on one another, as proposed by *e.g.*, Golubski & Abrams (2011); these are hypergraphs. Hypergraphs are useful when interactions rely, not only on species, but also on other species interactions: for example, an opportunistic pathogen may not be able to infect a healthy host, but may do so if the host's immune system is already being compromised by another infection. Hence it is not only species, but also their interactions, which interact. Such higher-order interactions can be detected through comparing observed species density or performance to that obtained under a dynamic model without higher-order interactions (Billick & Case 1994; Mayfield & Stouffer 2017). As using these concepts in ecological research represents a recent development, there is little methodology to describe systems represented as multigraphs or hypergraphs, and we will only mention them briefly here. In a way, methodological developments on these points are limited by the lack of data to explore their potential. As the interest among network ecologists will increase for systems in which the current paradigm of species–species interactions falls

short, we expect that the inflow of data will stimulate the emergence of novel methods.

Formally, all of these structures can be represented with the formalism of graph theory. A graph G is defined as an ordered pair (V, E), where every element of E (the edges) is a two-element subset of V (the nodes). From this simple structure, we can measure a large number of properties (see *e.g.*, Newman 2010 for an introduction). A simple graph contains neither self-edges (a node is linked to itself) or multiedges (the same two nodes are linked by more than one type of edge), whereas a multigraph contains at least one multiedge. As we illustrate in fig. 3.1, edges can be directed (*e.g.*, A eats B), or undirected (*e.g.*, A and B compete); unweighted (*e.g.*, A pollinates B) or weighted (*e.g.*, A contributes to 10% of B's pollination). In the context of studying ecological interactions, V is a set of ecological objects (taxonomic entities, or other relevant components of the environment), and E are the pairwise relationships between these objects. As both the strengths of interactions and their direction are highly relevant to ecological investigations, data on species interactions are most often represented as networks: directed and weighted graphs. We use network as a synonym for "graph" throughout. Species interaction networks can, finally, be represented as unipartite or bipartite networks. Unipartite networks are the more general case, in which any two nodes can be connected; for example, food webs or social networks are unipartite (Post 2002; Dunne 2006). Unipartite networks can represent interactions between multiple groups; for example, food webs can be decomposed into trophic levels, or trophic guilds. Bipartite networks, on the other hand, have nodes that can be divided in disjointed sets T (top) and B (bottom), such that every edge goes from a vertex from T, to a vertex from B; any ecological community with two discrete groups of organisms can be represented as a bipartite network (*e.g.* plant and mutualists, Jordano & Bascompte 2013; parasites and hosts, Poulin 2010; phage and bacteria, Weitz et al. 2013). It is possible to represent k-partite networks, *i.e.* networks with k discrete "levels". This formalism has been used for resources/consumers/predators (Chesson & Kuang 2008), and other plant-based communities (Fontaine et al. 2011a). Tripartite networks
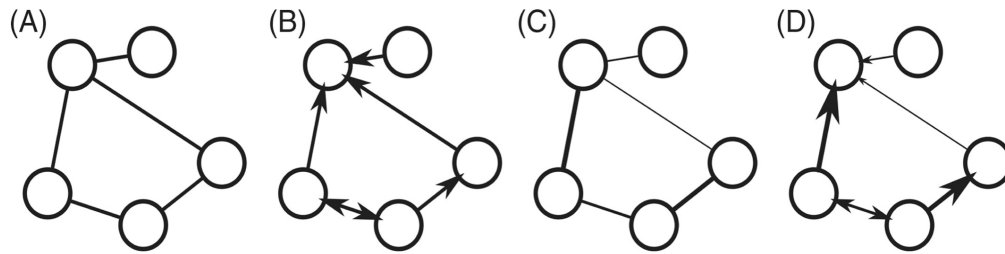
Figure 3.1 Differences between (un)weighted and (un)directed graphs. Graphs (A) and (C) are undirected, and graphs (A) and (B) are unweighted. Arrows thickness in graphs (C) and (D) represents the strength of the link.

are usually analyzed as collections of bipartite networks, or as unipartite networks. There still exists few data on ecological k-partite networks, and it is therefore difficult to establish solid recommendations about how they can be analyzed; this is a part of the field in which methodological developments are still needed and ongoing.

Networks can be represented using their adjacency matrix (**A**). For a unipartite network containing S species, **A** is a square matrix of dimensions (S, S). For a bipartite network containing $T + B$ species, the dimensions are (T, B), and the **A** matrix is usually referred to as the incidence matrix. In both cases, the elements a_{ij} of the matrix indicate whether species i interact with species j. In unweighted networks, $a_{ij} = 1$ when i and j interact and 0 otherwise. In weighted networks the strength of the interaction is given, instead of being set to unity. Note that in weighted networks, the strength of the interaction is not necessarily between 0 and 1; if the strength of interactions depicts the raw effect of one population on another, then it can take on both negative and positive values. The adjacency matrix is symmetrical for undirected networks, because $a_{ij} = a_{ji}$. In simple networks, the matrix diagonal is empty as there are no self-edges (which, ecologically, could represent autophagy, breastfeeding in mammals or cannibalism). We would like to note that **A** is not the de facto community matrix: in some situations, it can be more profitable to describe the community using its Jacobian matrix, *i.e.*, one in which a_{ij} represents the net effect of species i on species j (May 1972, 1974; Gravel et al. 2016b; Monteiro & Faria 2016; Novak et al. 2016).

3.3 What can we learn with ecological networks?

Here, unless otherwise stated, we will focus on describing measures of the structure of unweighted, directed networks (*i.e.*, either the interaction exists, or it does not; and we know in which direction it points), to the exclusion of quantitative measures that account for the strength of these interactions. In most cases, quantitative variations of the measures we present do exist (see *e.g.*, Bersier et al. 2002), and share a similar mathematical expression. We think that focusing on the simplifying (yet frequently used) unweighted versions allows one to develop a better understanding, or a better intuition, of what the measure can reveal. There is a long-standing dispute (Post 2002) among ecologists as to whether "arrows" in networks should represent biomass flow (*e.g.*, from the prey to the predator) or interaction (*e.g.*, from the predator to the prey). Because not all interactions involve biomass transfer, and because networks may be used to elucidate the nature of interactions, we will side with the latter convention. In general, we will assume that the interaction goes from the organism establishing it to the one receiving it (*e.g.*, from the pollinator to the plant, from the parasite to the host, etc.).

3.3.1 What do communities look like?

Order, size and density   During the last decades, various network measures have been developed to characterize the general structure of interacting communities, capturing both species identity and their interactions (Dunne et al. 2002b; Montoya et al. 2006; Allesina & Pascual 2007; Thompson et al. 2012). Most of these measures encompass and supplement usual measurements in community ecology. In addition to how many species there are, and which species are in local area, knowledge of their interactions is an additional layer of information that network measures exploit to quantify biodiversity.

A first descriptor of a network is its order (S), *i.e.*, the total number of nodes. If nodes are species, order measures the species richness of the community described by the network G.

The total number of interactions (L) is the size of the network. From these two measures is computed the linkage density $\frac{L}{S}$ (*e.g.*, Bartomeus 2013), which is the mean number of interactions per node – or simply, if a random species is selected, how many interactions it would be expected to have. Linkage density should be considered with caution as it can be misleading: the distribution of interactions among nodes in species interaction networks is rarely uniform or normal (Williams 2011), and a minority of species are known to establish a majority of interactions (Dunne et al. 2002a). Moreover L is known to scale with S^2 (Cohen & Briand 1984; Martinez 1992), at least in trophic interaction networks.

This observation that L scales with S^2 has cemented the use of an analog to linkage density, the connectance (Co), as a key descriptor of network structure (Martinez 1992). Connectance is defined as $\frac{L}{m}$, *i.e.*, the proportion of established interactions (L), relative to the possible number of interactions m. The value of m depends of the type of network considered. In a unipartite directed network, m is S^2. In a directed network in which species cannot interact with themselves, m is $S(S-1)$. In an undirected network, m is $S\frac{S-1}{2}$ if the species cannot interact with themselves, and $S\frac{S+1}{2}$ if they can. In a bipartite network, m is $T \times B$, the product of the number of species at each level. The connectance varies between 0 if the adjacency matrix is empty to 1 if its entirely filled. It is also a good estimate of a community sensitivity to perturbation (Dunne et al. 2002a; Montoya et al. 2006) as well as being broadly related to many aspects of community dynamics (Vieira & Almeida-Neto 2015). Although simple, connectance contains important information regarding how links within a network are distributed, in that many network properties are known to covary strongly with connectance (Poisot & Gravel 2014; Chagnon 2015), and the fact that most ecological networks "look the same" may be explained by the fact that they tend to exhibit similar connectances (fig. 3.2). Poisot & Gravel (2014) derived the minimum number of interactions that a network can have in order for all species to have at least one interaction. This allows us to express connectance in the $[0; 1]$ interval, where 0 indicates that the network has the least possible number of interactions.
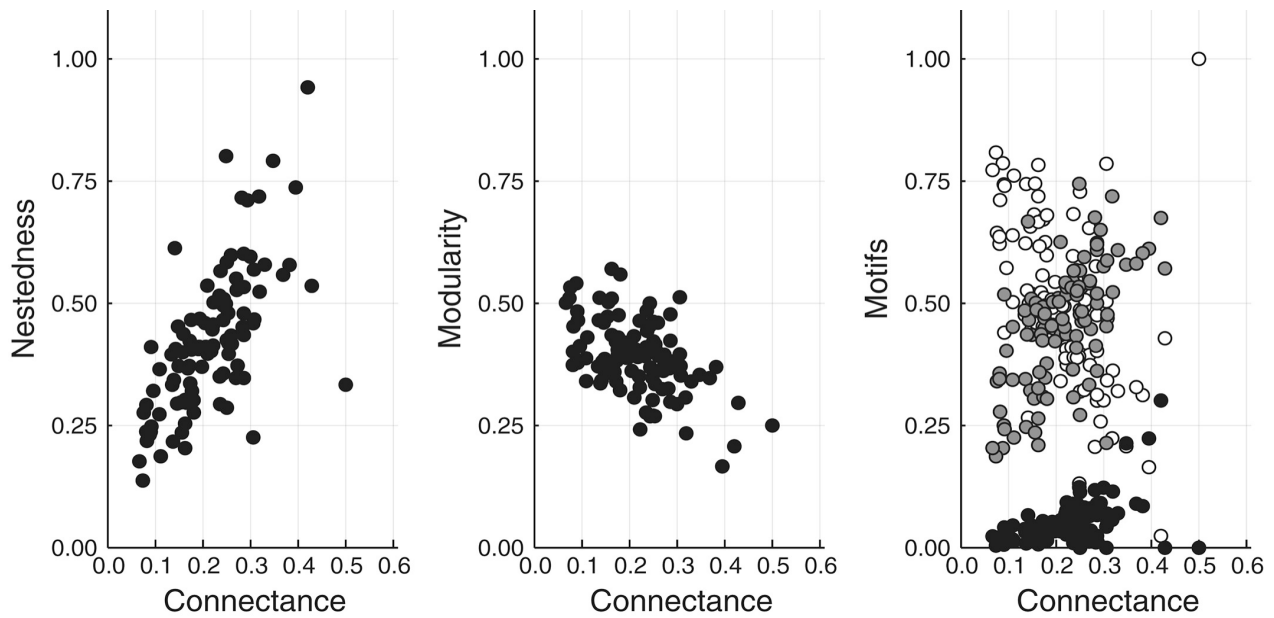
Figure 3.2 To illustrate the strong relationship betweeen connectance and other network measures, we measured the nestedness using η, modularity (best partition out of 100 runs), and the relative frequencies of three bipartite motifs (white, sparsely connected; grey, partially connected; black, fully connected) in 102 pollination networks. The sparsely connected motif represents two independent interactions. The partially connected motif represents the addition of one interaction to the sparsely connected one, and the fully connected motif includes the addition of another interaction. All of these measures have a strong covariance with connectance, and for this reason, the comparison of networks with different connectances must rely on randomizations. For data, methods, and code see **https://osf.io/82ypq/**.

Interactions repartition within the networks The majority of real-world species interaction networks are highly heterogeneous with regard to interactions distribution among nodes. It is possible to study the degree distribution of the network (the distribution of the number of interactions per node, see paragraph below). The way interactions are organized (distributed) among the nodes reflects ecological constraints and can be studied using various methods.Quantitative measures of different structures have been developed from graph theory and have played a growing role in understanding the evolution and functioning of ecological communities – in particular, because these measures add a small amount of information (comparatively to measures presented later below), they are a natural first step in moving away from a species-centric view of community into the arguably more realistic species-and-interactions view that networks capture well.

If the degree of a node is its number of interactions, then the degree distribution $P(k)$ measures the probability that a species has k interactions within the network. The degree distribution can be calculated as $P(k) = N(k)/S$ where $N(k)$ is the number of nodes with k interactions, and S is the total number of species in the network. The degree distribution allows identification of important nodes, such as potential keystone species (Solé & Montoya 2001 ; Dunne et al. 2002b), generalists, and specialist species (Memmott et al. 2004). In directed networks, the degree distribution can be divided into in-degree and out-degree. These respectively correspond to species vulnerability (*e.g.*, number of predators in food webs) and generality (*e.g.*, number of resources in food webs). It is often assumed that the distribution of degree in networks should resemble a power law (Strogatz 2001; Caldarelli 2007). In other words, the proportion P(k) of nodes with degree k should be proportional to $k^{-\gamma}$ (but see see Jordano et al. 2003 – a truncated power law may be a more accurate description). Assuming that power laws are an appropriate benchmark is equivalent to assuming that ecological networks are structured first and foremost by preferential attachment, and that deviation from power-law predictions suggests the action of other factors. Dunne et al. (2002a) found that, at least in

food webs, ecological networks tend not to be small-world or scale-free (*i.e.*, having a specific degree distribution; Caldarelli 2007), but deviate from these rules in small yet informative ways (specifically, about prey selection or predator avoidance). Opportunistic attachment and topological plasticity have been suggested as mechanisms that can move the system away from predictions based on power laws (Ramos-Jiliberto et al. 2012; Ponisio et al. 2017). We suggest that deviations from the power law be analysed as having intrinsic ecological meaning: why there are more, or fewer, species with a given frequency of interactions may reveal reasons for and/or constraints on particular species interactions.

The network diameter gives an idea of how quickly perturbations may spread by providing a measure of how dense the network is. Diameter is measured as the longest of all the shortest distances (d_{ij}) between every pair of nodes in the graph (Albert & Barabási 2002), where d_{ij} is the length of the shortest path (sequence of interactions) existing between the nodes i and j. A small diameter indicates the presence of a densely connected nodes, or hubs, hence fast propagation between nodes which may make the network more sensitive to perturbation (*e.g.* rapid spread of a disease; Minor et al. 2008). The diameter is relative to the number of nodes in the network, since it relies on counting the number of edges in a path, which may become larger as the network order increases. To overcome this issue, the diameter can also be measured as average of the distances between each pair of nodes in the network.

Aggregation of nodes based on their edges   From the heterogeneous repartition of interactions between nodes in species interaction networks, certain structures and groupings of interactions around nodes emerge. While the degree distribution hints at how edges are organized around single nodes, one can frame this question at the scale of the entire network. It is likely that other structures will appear when multiple nodes are considered at once. This can be done by analyzing what types of relationships the nodes (representing species, etc.) are typically embedded in (*e.g.*, competition, intraguild predation), through the analysis of motifs distribution, or by determining if there are nodes found in dense clusters or non-overlapping

compartments, forming modular communities.

Species interaction networks can be decomposed into smaller subgraphs of n species, called motifs (Milo et al. 2002). The smallest modules to which they can be decomposed are three-species motifs (Holt 1997). The relative frequencies of each of these motifs holds information about network structure. There are 13 possible three-nodes motifs in directed networks, each representing a different relationship between three nodes, such as competition between A and B for a shared resource C ($A \rightarrow C \leftarrow B$), or a linear chain between A, B and C ($A \rightarrow B \rightarrow C$). Among these 13 motifs, some are present in species interaction networks with a lower or higher frequency that what is expected in random networks. Motif distributions are characteristic of network type (neuronal, electrical, social, ecological, and so on). In food webs for example, motifs under- and over-representation has been found to be consistent across different habitats (Camacho et al. 2007; Stouffer et al. 2007; Borrelli 2015). In ecological networks, motifs have been referred to as the basic building blocks of communities, as they represent typical relationships between species. Studying their distribution (*i.e.*, how many of each type of motif there is this network) offers an opportunity to bridge the gap between two traditional approaches (Bascompte & Melián 2005), namely the study of the dynamics of simple modules such as omnivory or linear food chain (Pimm & Lawton 1978; Holt 1996; McCann et al. 1998), and the analysis of aggregated metrics describing the community as a whole. Motif distributions have been used to study the processes underlying the assembly and disassembly of ecological communities (Bastolla et al. 2009), as well as of the link between community's structure and dynamics (Stouffer & Bascompte 2011). More recently, motifs have also been used to define species' trophic roles in the context of their community (Baker et al. 2014) and to link this role to the network's stability (Borrelli 2015).

The clustering coefficient is useful to estimate the "cliquishness" of nodes in a graph (Watts & Strogatz 1998) – that is their grouping in closely connected subsets. It measures the degree to which the neighbours of a node are connected (the neighbourhood of a node i is composed

of all of the nodes that are directly connected to i). In other words, it gives an idea of how likely it is that two connected nodes are part of a larger highly connected group or "clique". Two different versions of the clustering coefficient (CC) exist. First, it can be defined locally, for each node i (Watts & Strogatz 1998). In this case $cc_i = \frac{2N_i}{k_i(k_i-1)}$ where k_i is i's degree (its number of neighbours) and N_i is the total number of interactions between i's neighbours. It describes the fraction of realized edges between i's neighbours and thus varies between 0 (none of i's neighbours are connected) and 1 (all of them are connected, forming a "clique"). From this measure, we can calculate the average local clustering coefficient: $CC_1 = \frac{\sum_i c_i}{S}$ where S is the total number of nodes. This first version describes the "cliquishness" of a typical neighbourhood, but has the drawback of giving more influence to nodes with a small degree. Nevertheless, the clustering coefficient provides a way of characterising the structure of the graph through the analysis of CC_k, which is the average of the cc_i of all nodes of degree k, and specifically of the distribution of CC_k across multiple values of k. The clustering coefficient can also be defined globally, for the entire graph (Soffer & Vazquez 2005; Saramäki et al. 2007) and is calculated as follows $CC_2 = \frac{3N_t}{N_c}$, where N_t is the number of triangles in graph G (a is connected to b and c, b to a and c and c to a and b) and N_c is the number of three-node subgraphs (e.g., a is connected to b and c, b and c are connected to a but not to each other). Kim (1993) suggested that this property of a network can be used to infer competition, but this has to our knowledge received little attention in ecology.

Whereas clustering analysis gives information about the grouping of nodes within their immediate neighbourhood (but no information about the identity of nodes in this neighbourhood), a measure of modularity gives a similar information at a larger scale (Gauzens et al. 2015). Network modularity measures how closely connected nodes are divided in modules, also called compartments (Olesen et al. 2007). A module is defined as a subsystem of non-overlapping and strongly interacting species (see fig. 3.3, matrices C and D for a comparison of the structures of modular and non-modular matrices). The modular structure of graphs has been studied because
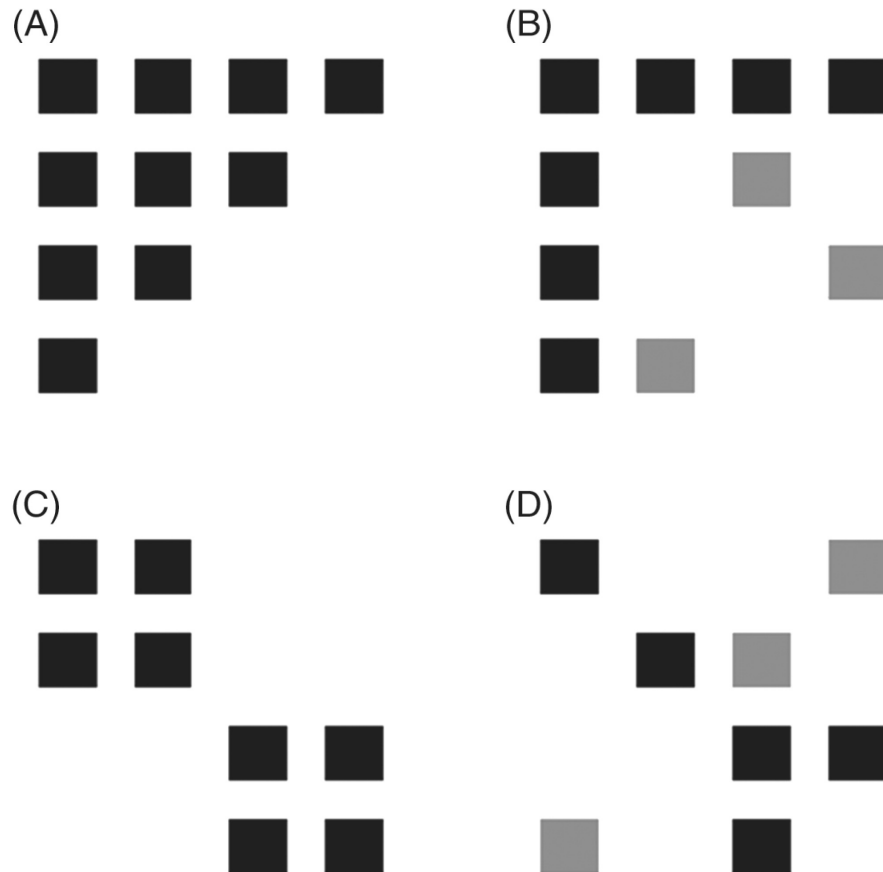
Figure 3.3 Illustration of the nested and modular structure of networks, represented as matrices. A is a perfectly nested matrix; in B, three interactions (in grey) have been displaced to lose the perfectly nested structure. C is a perfectly modular network; in D, three interactions have been displaced to lose the modular structure.

of its dynamical implications, in that modularity promotes stability by containing perturbations within a module, thereby constraining their spreading to the rest of the community (Stouffer & Bascompte 2010, 2011). This has been a key argument in the diversity-stability debate (Krause et al. 2003). A major challenge when studying the modularity of species interaction networks is to find the best subdivision of the network. Several methods have been developed for this purpose, which can be classified into three categories: i) classical optimization of a modularity function that maximizes link density within modules (Guimerà et al. 2004; Newman 2004; Newman & Girvan 2004; Guimerà & Amaral 2005; Guimerà & Nunes Amaral 2005), ii) probability mixture strategies (stochastic blockmodels, usually referred to as group model in ecology; Holland et al. 1983; Allesina & Pascual 2009a) and iii) modular flow analysis based on maps of random walk (Rosvall & Bergstrom 2008; Rosvall et al. 2009; Farage et al. 2020). These usually resulting in different groupings, reflecting the fact that there is not one true grouping of nodes in ecological networks. The method must thus be chosen carefully to fit the type of information one wants to reveal.

The optimization of a modularity function is by far the most popular in ecology The principle underlying this function is to find the optimal subdivision that maximizes the number of interactions within modules while minimizing the number of interactions between modules. The calculated modularity is then compared with a null model that has the same number of links and nodes, with the links connected to each other randomly. Modularity optimization has a resolution limit (in that its performance decreases with the size of the network) making it less reliable for large species interaction networks (Fortunato & Barthélemy 2007); there are methods designed specifically to work on thousands of nodes and more (see *e.g.* Liu & Murata 2009). To compare outcomes of different modularity measurements, it is possible to use an a posteriori method. In a network where modules are already found, the realized modularity (Q'_R) measures the proportion of interactions connecting nodes within modules (Poisot 2013).

This is expressed as

$$Q'_R = 2 \times \frac{W}{L} - 1 \,, \qquad\qquad (3.1)$$

where W is the number of interactions within modules, and L is the total number of interactions. This takes on a value of 1 when modules are disconnected from one another (which is not true of other modularity functions that account for the probability of establishing an edge). This measure can take on negative values if there are more interactions between modules than within them, which can be viewed as a non-relevant partitioning of the community.

Nestedness Species interaction networks can also present a nested structure (see fig. 3.3, matrices A and B for a comparison of the structures of nested versus non-nested matrices), where the species composition of small assemblages are subsets of larger assemblages. In food webs, a nested structure occurs when the diet of the specialist species is a subset of the diet of the more generalist species – and where the predators of species are nested as well. The analysis of nestedness has revealed ecological and evolutionary constraints on communities. For example, it has been hypothesized that a nested structure promotes greater diversity by minimizing competition among species in a community (Bastolla et al. 2009). Various metrics have been developed to quantify nestedness (Ulrich 2009; Ulrich et al. 2009). Most are based on the principle that when a matrix is ordered by rows and columns (that is descending in rank from above and from the left) a nested network will present a concentration of presence data in the top-left corner of the matrix, and a concentration of absence data in the opposite corner [see Staniczenko et al. (2013) for an exception; see fig 3.3C]. Numerous studies (Rodriguez-Girones & Santamaria 2006; Fortuna et al. 2010; Flores et al. 2011) use the proportion of unexpected presence or absence in the matrix to quantify nestedness. Seemingly the most widely used measure of nestedness is based on overlap and decreasing fills (NODF), as suggested by Almeida-Neto et al. (2007); Bastolla et al. (2009) suggested that η can complement NODF,

in that η does not require a re-ordering of the nodes (*i.e.*, there is no need to put the most densely connected nodes first, and the least densely connected nodes last). As per Bastolla et al. (2009), η is defined as:

$$\eta(\mathbf{A}) = \frac{\sum_{i<j} n_{ij}}{\sum_{i<j} \text{minimum}(n_i, n_j)} \tag{3.2}$$

where n_{ij} is the number of common interactions between species i and j, and n_i is the number of interactions of species i. Note that this formula gives the nestedness of rows with regard to the columns, though one can also measure the nestedness of columns with regard to rows as $\eta(\mathbf{A}')$, and calculate the nestedness of the whole system as the average of these two values. We suggest that, since it does not rely on species re-ordering, η can be used over NODF or other nestedness measures. There are some caveats to this argument, however. First, the number of permutations for NODF is known, and for species-poor networks, they can be computed in a reasonable time. Second, NODF can help understanding how different orderings of the matrix (*e.g.*, informed by species traits such as interaction strength or forbidden links)contributes to nestedness – if this is the question of interest, then NODF is the logical choice (Krishna et al. 2008). Once ordered by degree, NODF and η are identical (with the exception that NODF accounts for decreasing fill, whereas η does not). Finally, η has the undesirable property of always giving the same value depending only on the degree distribution. Therefore, any permutation of a network that maintains the degree distribution will give the same value of η, which greatly impedes hypothesis testing.

Intervality   A last measure of the structure of species interaction networks is their intervality. A network is "interval" when it can be fully explained by one dimension (trait). An interval food web with species ordered by their body mass, as an example, has predator eating a consecutive range of prey, that all fall into a range of body masses (Eklöf & Stouffer 2015), or are closely related from a phylogenetic standpoint (Eklöf & Stouffer 2015). The first step in calculating

intervality is to identify a common trait along which nodes can be ordered. This can be body mass in the case of food webs, but can also be a property derived from their position in the network, such as their degree; indeed, a nested bipartite network is interval when species are organized by decreasing degree. Intervality then measures how well interactions of all species can be described by this trait. Most unipartite ecological networks are close to being interval with one or several dimensions, such as defined by body size (Zook et al. 2011) or arbitrary traits derived from the interactions themselves (Eklöf et al. 2013b). There are several methods to quantify a network's intervality. Cattin et al. (2004a) quantified the "level of diet discontinuity" using two measures: (i) the proportion of triplets (three species matrix) with a discontinuous diet (*i.e.*, at least one species gap), in the whole food web (D_{diet}), and (ii) the number of chordless cycles (C_{y_4}). A cycle of four species (a graph cycle is a subset of species, here 4, that are connected by a continuous path of interactions such that the first species is also the last)is considered as chordless if at least two species out of the four are not sharing prey, so the diets cannot be totally interval. Nevertheless, these two measures only give a local estimation of intervality. Stouffer et al. (2006) proposed to measure the intervality of the entire network by re-organizing the interaction matrix to find the best arrangement with the fewest gaps in the network. This is a stochastic approach that by definition does not guarantee finding the global optimum, but has the benefit of working at the network scale rather than at the scale of triplets of species.

3.3.2 How are communities different?

Detecting spatial and temporal variation in ecological networks, and associating these variations with environmental factors, may yield insights into the underlying changes in ecosystem functions, emergent properties, and robustness to extinction and invasion (Tylianakis et al. 2007; Tylianakis & Binzer 2013). These efforts have been hindered by the difficulty of quantifying variation among interaction networks. The challenge lies in finding a meaningful

way to measure the dissimilarity between networks (Dale & Fortin 2010). Given the ecological properties or processes of interest, a direct comparison – not always computationally tractable – may not be necessary. Hence, networks can be indirectly compared through their properties (*e.g.*, degree distribution, connectance, nestedness, modularity, etc.). Multivariate analyses of network metrics have been used to estimate the level of similarity between different networks (Vermaat et al. 2009; Baiser et al. 2012), while null models were used to compare observed values statistically to their expected random counterparts (*e.g.*, Flores et al. 2011).

In the situation where several networks share a large enough number of species, one can alternatively compare how these shared species interact. This approach can be particularly useful along environmental gradients (Tylianakis et al. 2007; Tylianakis & Morris 2017). It represents a second 'dimension' of network variability, where in addition to changes in higher order structure, changes at the scale of species pairs within the networks are accounted for. This variation is more readily measured through a different approach to sampling, where instead of relying on the sampling of a large number of networks in different environments, efforts are focused on the same system at reduced spatial or temporal scales. The development of methods to analyse replicated networks is still hampered by the lack of such data; this is especially true in food webs. Replicated food webs based only on the knowledge of the local species and their potential interactions (*e.g.*, Havens 1992) are not always appropriate: by assuming that interactions always happen everywhere, we do not capture all sources of community variation (in addition to the issue of co-occurrence being increasingly unlikely when the number of species increases). Sampling of ecological networks should focus on the replicated documentation of interactions within the same species pool, and their variation in time and space (Poisot et al. 2012; Carstensen et al. 2014; Olito & Fox 2015), as opposed to relying on proxies such as comparison of different communities across space (Dalsgaard et al. 2013), or time (Roopnarine & Angielczyk 2012; Yeakel et al. 2014).

Analysis of network structure measures has so far played a central role in the comparison of

networks and in the search for general rules underpinning their organization (Dunne 2006; Fortuna et al. 2010). Notably, the number of species affects the number of interactions in real ecological networks (Martinez 1992; Brose et al. 2004), and thus many other network properties (Dunne 2006). Some measures of network structure covary with expected ecological properties, such as species abundance distributions (Blüthgen et al. 2008; Vázquez et al. 2012; Canard et al. 2014), network size and sampling intensity (Martinez et al. 1999; Banašek-Richter et al. 2004; Chacoff et al. 2012). This issue can seriously limit the interpretation of network measures and their use for network comparison. Furthermore, most of these measures are highly correlated among themselves: Vermaat et al. (2009) reported that network variation can be reduced largely along three major axes related to connectance, species richness (which is tied to connectance because the number of interactions scales with the number of species) and primary productivity (which is hard to measure, and is not easily defined for all systems). More recently, Poisot & Gravel (2014) and Chagnon (2015) showed that because of constraints introduced by the interaction between connectance and network size, the covariation of the simplest measures of network structure is expected to be very strong. As a consequence, it is barely possible to make robust network comparisons using the variations in these basic descriptors. We therefore need to go beyond these global network properties, and find meaningful alternatives that allow a better understanding of the ecological differences between networks.

Differences in global structure   Other methods accounting for the structure of the entire network have been developed. For example, some methods are based on the frequency distribution of small subnetworks including network motifs (Milo et al. 2002) and graphlets (a more general definition of motifs; Pržulj 2007; Yavero lu et al. 2015). The method of graph edit distance gives edit costs (each modification to the graph counts for one unit of distance) for relabelling nodes, as well as insertion and deletion of both nodes and interactions (Sanfeliu & Fu 1983), and therefore provides a well-defined way of measuring the similarity of two networks (this method has not been widely used in ecology). Other suitable measures

to determine network similarity are based on graph spectra (Wilson & Zhu 2008; Stumpf et al. 2012). Spectral graph theory (which is yet to be applied comprehensively to the study of species interaction networks, but see Lemos-Costa et al. (2015)) characterizes the structural properties of graphs using the eigenvectors and eigenvalues of the adjacency matrix or the closely related Laplacian matrix (the Laplacian matrix, defined as $\mathbf{D} - \mathbf{A}$, wherein $\mathbf{D}$ is a matrix filled with 0's in the off-diagonal elements, and the degree of each node is on the diagonal and accounts both for network structure and for degree distribution). Some methods allow the algorithmic comparison of multiple networks in which no species are found in common (Faust & Skvoretz 2002; Dale & Fortin 2010), and are primarily concerned with the overall statistical, as opposed to ecological, properties of networks.

Ecological similarity and pairwise differences   All of the aforementioned methods are adapted from other fields (usually physics) and focus on networks as mathematical abstractions. Developing new methods, rooted in ecological processes, to compare ecological networks would potentially provide new important insights. Poisot et al. (2012) presented a framework for measurement of pairwise network dissimilarity, accounting both for species and interactions turnover through space, time or along environmental gradients. This method extends the notion of β-diversity to the network of interaction underlying communities Following Koleff et al. (2003), this approach partitions interactions in three sets: shared by both networks, unique to network 1, and unique to network 2. The β-diversity can be measured by comparing the number of interactions shared and unshared by these three sets to reflect symmetry of change, gain/loss measures, nestedness of interaction turnover, etc. This method of network β-diversity can also be extended to multiple network comparisons using their relative difference from the same meta-network. While many measures of β-diversity exist to analyse compositional data, there is still a lack of a comprehensive methodology regarding their application to networks. A large part of this stems from the fact that species interactions require the species pair to be shared by both communities, and consequently some analyses require that the species pair is

shared by two communities: measures of network β-diversity are strongly constrained by the structure of species co-occurrence. If no species pairs co-occur, or if no two networks have common species, these methods cannot give informative results (the dissimilarity being, by default, complete) – as of now, this suggests that a tighter integration of these methods with research on compositional turnover is needed, especially to understand the threshold of shared species below which they should not be applied. In addition, none of the current methods seem sufficient to characterize the structure for a meaningful comparison and to extract information hidden in the topology of networks (as they ignore network-level structure, *i.e.*, emerging from more than direct interactions), and the development of future methods that work regardless of species composition seems like a straightforward high-priority topic. Finally, this framework would benefit from a better integration with quantitative measures. Using Bray-Curtis (or equivalent) measures to assess difference between networks for which interaction strengths are known would allow us to quantify dissimilarity beyond the presence or absence of interactions.

3.3.3 What do species do?

Not all species in large communities fulfill the same ecological role, or are equally important for processes and properties acting in these communities. As species interactions are a backbone for fundamental mechanisms such as transfer of information and biomass, one can expect that the role of a species reflects its position within its community, organized by trophic level, abundance, body size or other ecologically meaningful organizing principles. In species interaction networks, it is possible to measure the position and the role of species in different ways, giving different ecological information.

Centrality   Centrality is a measure of how "influential" a species is, under various definitions of "influence". It has been used to identify possible keystone species in ecological networks (Jordán & Scheuring 2004; Martín González et al. 2010). We note that the ability of network structure measures to identify keystone species is highly dubious; the canonical definition of a

	Betw.	Eigen.	Degree	Close.
(light grey)	**0.36**	0.15	0.12	**0.18**
(dark grey)		0.13	0.12	0.12
(white)	0.32	**0.17**	**0.19**	0.17
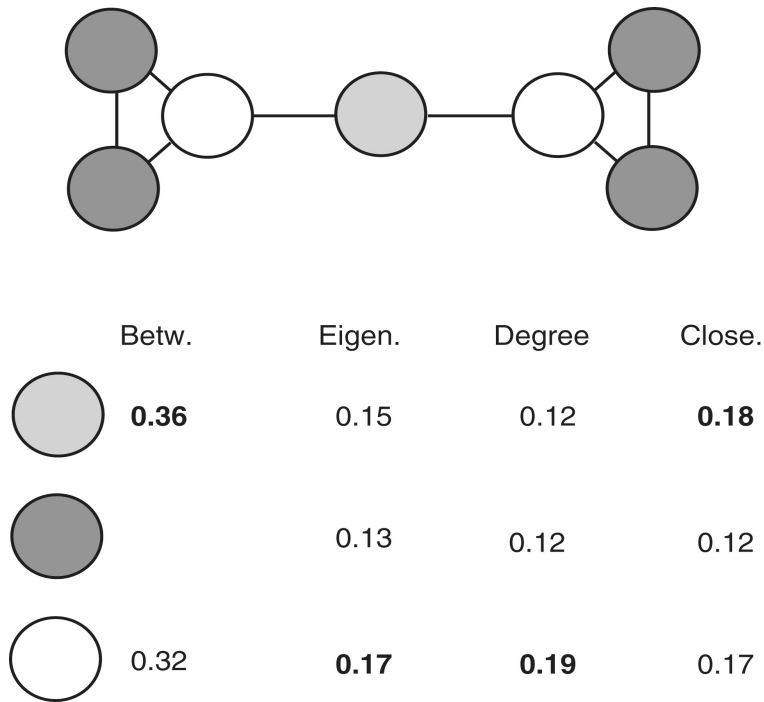
Figure 3.4 On the simple graph depicted at the top (nodes of the same shade have the same centralities), we measured centrality using betweenness, eigen centrality, degree centrality, and closeness. The values have been corrected to sum to unity. The value in bold gives the most central family of nodes for the given measure. This example illustrates that different measures make different assumptions about what being "central" means. The dark-grey nodes do not have a betweenness centrality value; some software returns 0 for this situation.

keystone species (Paine 1969) requires knowledge about biomass and effects of removal, which are often not available for network data, and makes predictions that are primarily about species occurrences. These measures may be able to identify list of candidate keystone species, but this requires careful experimental/observational validation. Nevertheless, knowledge of network structure allows us to partition out the effect of every species in the network. For example, in networks with a nested structure, the core of generalist species have higher centrality scores, and the nested structure thought to play an important role for network functioning and robustness (Bascompte et al. 2003). We provide an illustration of some centrality measures in fig. 3.4.

Degree centrality ($C_D(i) = k_i$; Freeman (1977)) is a simple count of the number of interactions established by a species. In directed networks, this measure can be partitioned between in-degree (interactions from others to i) and out-degree (interaction from i to other).

It is a local measure, that quantifies the immediate influence between nodes. As an example, in the case of a disease, a node with more interactions will be more likely both to be infected and to contaminate more individuals (Bell et al. 1999). To compare species' centrality, C_D has to be normalized by the maximum degree ($\langle C_D \rangle = C_D/k_{max}$).

Closeness centrality (C_C) (Freeman 1978; Freeman et al. 1979) measures the proximity of a species to all other species in the network, and is therefore global in that, although defined at the species level, it accounts for the structure of the entire network. It is based on the shortest path length between pairs of species and thus indicates how rapidly/efficiently a node is likely to influence the overall network. The node with the highest C_C is closer to all other nodes than any other nodes and will thus affect more rapidly the overall network if, for example, there is a perturbation (Estrada & Bodin 2008). Formally, C_C is defined as

$$C_C(i) = \sum_{j \neq i} \frac{n-1}{d_{ji}}, \tag{3.3}$$

where d_{ij} is the shortest path length between i and j, and n is the number of species.

Betweenness Centrality (C_B) (Freeman 1977) describes the number of times a species is between a pair of other species, $i.e.$, how many paths (either directed or not) go through it. This measure is thus ideal to study the influence of species loss on fragmentation processes, for example (Earn 2000; Chadès et al. 2011; McDonald-Madden et al. 2016). Nodes with high C_B values are considered as module connectors in the network. The value of C_B is usually normalized by the number of pairs of species in the network excluding the species under focus, and is measured as

$$C_B(i) = 2 \times \sum_{j<k; i \neq j} \frac{g_{jk}(i)/g_{jk}}{(n-1)(n-2)}, \tag{3.4}$$

where g_{jk} is the number of paths between j and k, while $g_{jk}(i)$ is the number of these paths that include i.

Eigenvector centrality (C_E – Bonacich 1987) is akin to a simulation of flow across interactions, in which each species influences all of its partners simultaneously. It then measures the relative importance of species by assigning them a score on the basis that an interaction with more influential species contributes more to a species' score than the same interaction with a low-scoring species (Allesina & Pascual 2009b). From a graph adjacency matrix $\mathbf{A}$, the eigenvector centrality of species i is given by

$$C_E(i) = \frac{1}{\lambda} \sum_j \mathbf{A}_{ij} C_E(j) , \tag{3.5}$$

where $\mathbf{A}_{ij}$ is 1 if i interacts with j and 0 otherwise, and λ is a constant. This can be rewritten as the eigenvector equation:

$$\mathbf{Ac} = \lambda \mathbf{c} , \tag{3.6}$$

wherein $\mathbf{c}$ is the vector of all values of C_E. As all values of C_E have to be positive, as per the Perron-Frobenius theorem, λ is the greatest eigenvalue of $\mathbf{A}$.

Finally, Katz's centrality (C_K – Katz 1953) is a measure of the influence of a node in the network. This measure takes into account all the interactions connecting a node to its neighbourhood. However, an immediate neighbour has more weight than a distant one. C_K is defined as

$$C_K(i) = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \alpha^k \mathbf{A}_{ij}^k , \tag{3.7}$$

wherein α is the attenuation constant, and k is the length of the paths between i and j. The α value is between 0 and $1/\lambda$, where λ is the largest eigenvalue of $\mathbf{A}$. Larger values of α give more importance to distant connections, thus allowing this measure to function either locally (immediate neighborhood) or globally (entire graph). C_K can be used in directed acyclic graphs (*e.g.*, trees), which is not true of C_E. This is also the only measure to have a

probabilistic equivalent (Poisot et al. 2016a).

Studying different measures of centrality provides important information regarding the roles of certain species/nodes. As an example, a species may have a low C_D and a high C_B, meaning that it plays a key role in connecting species that would not be connected otherwise even if it does not interact with them directly. A low C_D and a high C_C means that the species has a key role by interacting with important species. Because the absolute values of centrality vary with network size and connectance, Freeman et al. (1979) suggested that the centralization measure, rarely applied in ecology, be used when comparing centrality across networks. Centralization is defined, for any centrality measure C_x, as the sum of the differences between each node's centrality, and the highest centrality value ($\sum_i (C_x(i) - \max(C_x))$). This measure is then divided by the maximal possible value of centralization for a network with the same number of nodes and interactions, which in turn depends on the formulae used to measure centrality, and can be estimated based on random draws of the networks.

Species roles in the network   Species functional roles can be reflected in the interactions they establish (Coux et al. 2016), providing a clear bridge between network approaches and functional ecology studies. Functional traits are known to be correlated with the position of species in the network, either because they intervene directly in the interaction (Brose et al. 2006a; Alexander et al. 2013), constraining the set of possible interactions or their frequency, or because phenological incompatibilities prevent the interaction from happening (Olesen et al. 2011). For instance, (Petchey et al. 2008a) used allometric scaling of body size and foraging behaviour of individual consumers to predict species interactions. Scaling up to multiple traits, one can group species into functional clusters, based on their similarity. The distribution of some species-level network measures (*e.g.*, centrality, degree) can then be compared within and across groups (Petchey & Gaston 2002). This method usually does not account directly for interactions between species (Petchey et al. 2008a) but is useful when studying a process for which the influential traits are known, or to test the importance of a particular (complex of)

traits on a function. Moreover, when the trait used is correlated to diet choice (*e.g.*, body mass), and because we know that networks are usually interval(see paragraph on intervality), this may group interacting species.Note that one can, in this situation, adopt a very generous definition of what constitutes a trait: spatial grouping of species (Baskerville et al. 2011) is one example in which examining interactions in the light of species attributes provides ecological insights.

If external information on species traits is absent, the role of a species can be approached through the interactions it establishes within the network: species with similar interactions are often grouped into trophic species, and these can be assumed to have similar traits or lifestyles (this approach has mostly been used in food webs). Indeed, many food-web models (Williams & Martinez 2000; Cattin et al. 2004a) predict interactions between trophic groups, and not between species. Lumping species within trophic groups maintains the heterogeneity of interactions across groups, but removes all variability of interactions between species within the groups. As a consequence, species that bring unique interactions to a trophic group may be overlooked. Dalla Riva & Stouffer (2015) suggested an alternative to this approach: species positions are analysed before clustering them into groups (*i.e.*, there is a measure of position for every species), allowing explicit investigation of species interactions while avoiding obfuscation of the variance within groups.

Coux et al. (2016) measured the functional role of species, by applying functional dispersion $FDis$ (Laliberté & Legendre 2010) to the adjacency or incidence matrix of the network. Under this framework, as in Mouillot et al. (2013), the uniqueness of a species is hinted at by its distance to the centroid of all other species. We argue that this approach should be questioned for two reasons. First, it is sensitive to the ordination choices made. Second, it is not clear how it allows the comparison of results across different networks: not only does the position of a species vary in relation to other species in the network, it varies from one network to another. Note that centrality measures are not necessarily better at identifying which species are unique:

as we show in fig. 3.4, for some measures, non-unique nodes have high centrality values. We argue that the development of measures for node uniqueness should receive increased attention. In particular, measures that rely on ordination only account for first-order interactions, *i.e.*, direct interactions between species. As a consequence, a large part of the network structure, which emerges through consideration of longer chains of interactions, is not accessible via these methods.

Looking at network motifs is a promising way to address species functional roles and node uniqueness. Motifs are all the possible ways a fixed number of species (usually three or four) can interact. Within these motifs, species can occupy a variety of unique positions; for example, within a linear food chain, there are three distinct positions (bottom, middle, top), whereas a trophic loop has a single unique position. Within motifs with three species, 30 unique positions can be identified (Stouffer et al. 2012), and for each species, its frequency of appearance at each of these positions within networks has been shown to be an inherent characteristic conserved through its evolutionary history. This method has the advantage of grouping species that may be different in terms of guild or partners, but that contribute in the same way to the structure of the community. Based on this vector it is possible to identify species statistically that exhibit similar profiles. Motif positions tend to be well conserved both in time (Stouffer et al. 2012) and space (Baker et al. 2014), making them ideal candidates to be investigated alongside functional traits and phylogenetic history.

Partition based on modularity In large communities, some species are organized in modules (see "What do communities look like" part "Edges repartition within the graph"), within which they interact more frequently among themselves than with species of the same overall network but outside of their module. Guimerà & Nunes Amaral (2005) proposed that when functional or topological modules can be found in large networks, the functional role of a species can be defined by how its interactions are distributed within its module and with other modules. To identify these roles, the first step is to identify the functional modules of a large network (see

"What do communities look like" part "Edges repartition within the network"). The profile of species interactions is determined by using two measures.

First, the z-score quantifies how well-connected a species i is within its module m.

$$z_i = \frac{K_i - \overline{K}_{m_i}}{\sigma_{K_{m_i}}}, \tag{3.8}$$

where K_i is the degree of i within its module m_i; $\overline{K}_{m_i}$ is the average of K over all species of m_i and $\sigma_{K_{m_i}}$ is the standard deviation of K in m_i.

Second, the participation coefficient (PC) describes the profile of i's interaction with species found outside of the module m.

$$PC_i = \sum_{m=1}^{N_M} (\frac{K_{is}}{k_i})^2, \tag{3.9}$$

where k_i is the total degree of species i, meaning a count of all its connections, inter- and intra module. The PC of a species therefore varies between 0 (all interactions are within the module) and 1 (all interactions are uniformly distributed among all modules). The use of these indices is based on the assumption that species with similar interactions have similar traits and so are expected to play the same functional role.

Olesen et al. (2007) used these two values to divide species into four groups, based on a cutoff for z (2.5) and for PC (0.62). Species with low z and low PC are "peripherals" – they are not well connected within or between modules. Species with low z and high PC connect well between, but not within, modules, and are "connectors". Species with high z and low PC are "module hubs", well connected within their own modules but not with the outside. Finally, species with high z and high PC are "network hubs", connecting the entire community. In their analysis of plants and pollinators, Olesen et al. (2007) revealed that pollinators tend not to be module hubs, and are also less frequently network hubs than plants are.

Contribution to network properties. As species make differential contributions to network structure and processes, the removal of certain species will therefore have a greater effect on the community's stability and functioning, and these species are therefore stronger contributors to these processes. Differential contribution to several processes can be estimated in multiple ways: by performing removal/addition experiments in real ecological systems (*e.g.*, Cedar creek or BIODEPTH experiments), by analyzing the effect of a species extinction within empirical (Estrada & Bodin 2008) or simulated (Berlow et al. 2009) systems, by using a modelling approach and simulating extinctions (Memmott et al. 2007), or by analyzing the statistical correlation between an ecosystem property and species functional roles (Thompson et al. 2012). Another way to quantify the contribution of a species to a property P is to compare it to its contribution to the same property when its interactions are randomized (Bastolla et al. 2009). This method allows studying the contribution of a species' interactions, as the variation of interactions is intuitively expected to be faster than the variation of species. Indeed, because interactions require species to co-occur, because there are far more interactions than species, and because interactions have dynamics of their own, whether there will be more signal in interactions than in species presence is an hypothesis that should be tested on empirical systems in priority.

The contribution of a species to a given network measure after its interactions are randomized is

$$c_i = \frac{(P - \langle P_i^\star \rangle)}{\sigma_{P_i^\star}} \, , \tag{3.10}$$

where P is the property (nestedness, modularity, productivity, ...), $\langle P_i^\star \rangle$ and $\sigma_{P_i^\star}$ are the average and standard deviation, respectively, of the property across a set of random replicates for which species i interactions have been randomized. The effects of several traits or structural properties of species (such as centrality or species trophic roles) on their contributions to given measure can then be analyzed.

3.3.4 How similar are species interactions?

Some species exhibit a much larger set of interactions than others or form denser clusters within the network. One of the many challenges of ecology is to understand the causes and consequences of such heterogeneous species interactions. Species are, first and foremost, related by their phylogenetic history. We will not address this aspect here, because it does not easily integrate with network theory. We encourage readers to refer to Cadotte & Davies (2016) instead.

One way in which the heterogeneity of species interactions is quantified is through analysis of the overlap in their partners, known as ecological similarity. For simplicity, we will use the vocabulary derived from trophic networks, but these methods can also be applied to other types of ecological networks. Ecological similarity between species is a widely used concept that quantifies the resemblance between two species or "biotic interaction milieu" (McGill et al. 2006) and allows analyzing processes ranging from species niche (Elton 1927) and community assembly (Piechnik et al. 2008; Morlon et al. 2014) to trophic diversity (Petchey & Gaston 2002). The simplest and most widely used measure of pairwise ecological similarity is the Jaccard coefficient (Legendre & Legendre 2012):

$$S_J = \frac{a}{a + b + c} \tag{3.11}$$

where a is the number of shared partners, b the number of species that interact with only the first species and c with only the second species (for variations, see (Legendre & Legendre 2012)). The Jaccard similarity coefficient is widely used to estimate ecological similarity and competition between species (Rezende et al. 2009) but does not account for the shared absence of interactions (but see Chao et al. 2005). This is not a severe issue, as ecological networks tend to be extremely sparse, and therefore shared absence of interactions may not be informative. The similarity index has to be chosen with care depending on the focus of the study. In the

general equation above, consumers and resources are seen as perfectly equivalent (additively), but, in directed networks, it can be adapted to include consumer and resources as different dimensions of trophic activities and/or for dynamical food webs by including information about flows (Yodzis & Innes 1992). Once a similarity matrix is formed from all pairwise measurements, a hierarchical clustering can be performed to build a dendrogram, which gives information about the trophic diversity of species within a community and the relative uniqueness of species (but see Petchey et al. 2008b). Cophenetic correlation (Sokal & Rohlf 1962) can then be used to analyze how well several dendrograms, built using different methods, preserve the similarity between species (Yodzis & Winemiller 1999). The similarity of overall communities can also be estimated to see how similar, or dissimilar, species within it are when compared to null models (Morlon et al. 2014). For this purpose, the mean or maximum pairwise similarity is averaged across the whole network under consideration.

3.3.5   Is any of this significant?

Most network properties tend to be colinear (Vermaat et al. 2009), specifically because they covary with the number of species and links (MacDonald et al. 2020). For example, the number of interactions in a network with a known number of species will limit the possible values of nestedness, modularity, and so on (Poisot & Gravel 2014). As such, the value of any measure of network structure often needs to be compared to a range of possible values under a null model. The purpose of the null model is to search the null space of possible randomized networks (Fortuna et al. 2010), in a way that would yield an unbiased distribution of the measure of interest, to which the observed value is then compared. In practice, this approach is constrained by (i) the size of the null space to search, and specifically the fact that it varies with connectance (Poisot & Gravel 2014), and (ii) the computational burden of a thorough null space exploration.

 A large number of studies use the null hypothesis significance testing (NHST) paradigm to

assess the significance of an observed value of network structure. NHST works by generating randomized networks under a variety of constraints, measuring the property of interest on these randomizations, then commonly using a one-sample t-test with the value of the empirical measure as its reference. This is justified because, through the mean value theorem, the application of enough randomizations should yield a normal distribution of the simulated network measure (see Flores et al. 2011). Bascompte et al. (2003) used a probabilistic sampling approach, where the probability of drawing an interaction depends on the relative degree of the species; Fortuna & Bascompte (2006) used the same approach, with the distinction that all interactions have the same probability (equal to connectance). Drawing from a probability distribution in this manner has a number of shortcomings, notably the fact that some species can end up having no interactions, thus changing the network size (which Fortuna et al. 2010 termed "degenerate matrices"). An alternative approach is to use constrained permutations, where pairs of interactions are swapped to keep some quantity (the overall number of interactions, the degree of all species, and so on) constant. This approach is used in null models for species occupancy (Gotelli 2000; Gotelli & Entsminger 2003; Ulrich & Gotelli 2007). Stouffer et al. (2007) used an intermediate approach, where swapping was done as part as a "simulated annealing routine", to give the algorithm enough leeway to explore non-optimal solutions before converging (as opposed to just swapping, which has no definition of the optimality of a solution). Another possibility is to use alternatives to null model testing, as proposed by MacDonald et al. (2020). Seeing that the number of links may be best viewed not as a fixed but rather a probabilistic quantity (Poisot et al. 2016a), they suggest using mathematical models instead of simulation of random matrices to provide a domain of expected values. This type of alternative methods, as null models, requires more development. As of now, there are no clear recommendations as to which approach to sample the null space is the most efficient (or computationally feasible for large network sets), emphasizing the need for a more exhaustive comparison of the behaviour of these methods.

Hypotheses underpinning null models The most frequently used null models are topological, *i.e.*, they can search the null space based only on the matrix, and do not rely on ecological processes to generate random networks. We will focus on the subset of null models which generate a probability of observing an interaction based on different aspects of network structure; these probabilistic networks can be analysed directly or, as is most commonly done, converted into binary networks through random draws. There are three broad categories of null models (commonly used for bipartite networks) — based on connectance, based on degree distribution, and based on marginal degree distribution. Each family embodies a specific hypothesis about the sources of bias on the measured property. Type I (Fortuna & Bascompte 2006) null models are focused on connectance, where the probability of any two species i and j interacting is fixed as

$$P_{i \to j} = \frac{|E|}{|T| \times |B|} \, ,$$ (3.12)

where T and B are nodes from the "top" ($T = \{v \in V, k_{in}(v) = 0\}$) and "bottom" ($B = \{v \in V, k_{out}(v) = 0\}$) levels of the network (these methods were originally applied to bipartite networks). This model assumes that interactions are distributed at random between all species, without considering the degree of the species. Deviation from the predictions of this model indicate that the network measure of interest cannot be predicted by connectance alone.

Type II null models (Bascompte et al. 2003) add an additional level of constraint, in that they respect the degree distribution of the network (in degree k_{in} ; out-degree k_{out}). In a Type II network,

$$P_{i \to j} = \frac{1}{2} \left(\frac{k_{in}(j)}{|T|} + \frac{k_{out}(i)}{|B|} \right) \, ,$$ (3.13)

meaning that the interaction is assigned under the hypothesis that *i* distributes its outgoing interactions at random, and *j* receives its incoming interactions at random as well. In this

78

model, species with more interactions have a higher probability of receiving interactions in the simulated network. This conserves both the distribution of generality and vulnerability. Deviation from the predictions of this model indicate that the network measure of interest cannot be predicted by the degree distribution alone.

Finally, Type III models account for only one side of the degree distribution, and can be defined as Type III in, wherein

$$P_{i \to j} = \frac{k_{\text{in}}(j)}{|T|} \, , \qquad (3.14)$$

and Type III out, wherein

$$P_{i \to j} = \frac{k_{\text{out}}(i)}{|B|} \, . \qquad (3.15)$$

Deviation from the predictions of this model indicate that the network measure of interest cannot be predicted by the marginal degree distributions alone. Ecologically speaking, deviation from this null model means that the way interactions are established/received is sufficient to explain the observed structure. These models can be expressed in a sort of hierarchy. Type I introduces the least hypotheses, and should be applied first. If there is no significant deviation, then Type III models can be applied, then Type II. This approach has the important benefit of, in addition to determining which properties show a difference from the random expectation, giving insights about which aspect of the structure are responsible for this difference.

Topological and generative models It is important to note that these models, based on permutations, are purely topological. There is no difference, when deciding if an interaction should be assigned between two species, between *e.g.*, a plant-pollinator network, or a host-parasite network. One may want to test deviation from a null distribution that would be informed by ecological processes. To inject some processes into the null models used, several "generative" models have been proposed. In contrast to topological models, generative models

use core assumptions about ecological mechanisms to generate networks that mimic aspects of a template network. Arguably the most influential (despite it being limited to trophic interactions) is the "niche model" (Williams & Martinez 2000), that generates networks of trophic groups based on the hypothesis that feeding interactions are determined by an arbitrary niche-forming axis generally accepted or implied to be body-size ratios (Brose et al. 2006a). Gravel et al. (2013) showed that the parameters of this model can be derived from empirical observations. The niche model assumes a beta distribution of fundamental niche breadth in the entire network (in cases where the trait space is bound between 0 and 1); this assumption, close though it may be to empirical data, nevertheless has no mechanistic or theoretical support behind it. This suggests that so-called generative models may or may not be adequately grounded in ecological mechanisms, which implies the need for additional developments. Similar models include the cascade model and the nested-hierarchy model, but these tend to generate networks that are qualitatively similar to those of the niche model (Brose et al. 2006b). More recently, several models suggested that species traits can be used to approximate the structure of networks (Santamaría & Rodríguez-Gironés 2007; Bartomeus 2013; Crea et al. 2015; Olito & Fox 2015; Bartomeus et al. 2016). Finally, networks tend to be well described only by the structure of species abundances. Both in food webs (Canard et al. 2012) and host-parasite bipartite networks (Canard et al. 2014), modelling the probability of an interaction as the product of relative abundance appears sufficient to generate realistic networks. These generative models represent an invaluable tool, in that they allow building on mechanisms (although, as we illustrate with the niche model, not necessarily ecological ones) instead of observed relationships to generate the random expectations. The NHST-based analyses then proceed as with topological models, *i.e.*, the observed value is compared to the distribution of values in the theoretical space.

3.3.6 Future methods for novel questions

Surveying the methodological toolkit available to analyze ecological networks highlights areas in which future developments are needed. We identified, in particular, four topics that require additional attention.

Multi/hyper graphs Most of the tools to analyse species interaction networks are limited to node-to-node interactions, to the exclusion of node-to-interaction or interaction-to-interaction interactions. This limits the variety of biological situations that can be represented. Golubski & Abrams (2011) presented a number of situations that elude description in this way. For example, opportunistic infection by a pathogen O requires the pre-existence of an interaction between a pathogen P and an host H. This situation is better captured as (i) the existence of an interaction between H and P (noted L_{HP}) and (ii) the existence of an interaction between O and this interaction, noted $O \rightarrow L_{HP}$. Another hard-to-represent scenario is niche pre-emption: if a host H can be infected by either pathogen P_1 or P_2, but not both at the same time, then the interactions L_{HP_1} and L_{HP_2} interact antagonistically. This is a different situation from simple competition between P_1 and P_2. Although these are extremely important drivers of, for example, species distributions (Araújo & Rozenfeld 2014; Blois et al. 2014), the current methodological framework of ecological network analysis is not well prepared to represent these data.

External information Building on the basis suggested by Poisot et al. (2015c), Bartomeus et al. (2016) proposed that the mechanisms determining ecological interactions can be identified within a cohesive statistical framework, regardless of the type of ecological interaction. At its core, their framework assumes that interactions are the consequence of matching rules, *i.e.*, relationships between trait values and distributions. For example, a pollinator can get access to nectar if its proboscis is of a length compatible with the depth of the flower. Rather than

relying on natural history, these "linkage rules" (Bartomeus 2013) can be uncovered statistically, by modelling an interaction L_{ij} as a function $f(x_i, y_j)$ of the traits involved, wherein x_i and y_j are sets of traits for species i and j respectively. Procedures akin to variable selection will identify the traits involved in the interaction, and model selection can identify the shape of the relationship between trait values and interactions. There are two reasons for which this work is an important milestone in the modern analysis of ecological networks. First, it places interactions within the context of community ecology, by showing how they build upon, and influence, trait distributions. In particular, it draws attention to how structure of networks results both from the linkage rules and from the distribution of traits in the locality where the network is measured (Gravel et al. 2016a). Second, it does away with the necessity of topological models to generate random networks: identifying matching rules is the only step needed to generate random networks based on functional, biological hypotheses, thereby solving some of the concerns we identified with generative null models. We argue that this approach should be expanded to accommodate, *e.g.*, phylogenetic relationships among species. The ideal framework to study networks, and the one we should strive for, avoids considering interactions in isolation from other aspects of community structure – instead, it is explicit about the fact that none of these aspects are independent. Although this will come with additional mathematical and statistical complexity, this cost will be more than offset by the quality and the refinement of the predictions we will make.

Although documenting species, traits, and interactions seems like a daunting effort, there are novel approaches to accelerate the generation of data in some systems. For example, Bahlai & Landis (2016) showed that passive measurement based on citizen science (using Google Images) allows users to accurately document phenological matches and species interactions between flowers and bumblebees. Similarly, Evans et al. (2016) showed that sequencing of diet gives access to phylogenetic and interaction history within a single experiment. Addressing novel questions will require a diversification of the methodological toolkit of network ecologists,

as well as an improved dialogue between empiricists and theoreticians.

Networks of networks   An additional frontier for methodological development has to do with the fact that networks can be nested. A network of species–species interactions is the addition of interactions at the population level (Poisot et al. 2015c), themselves being aggregates of interactions at the individual level (Dupont et al. 2011, 2014; Melián et al. 2014). This is also true when moving from single-site to multi-site network analysis (Poisot et al. 2012; Canard et al. 2014; Carstensen et al. 2014; Trøjelsgaard et al. 2015). Local interaction networks exist in meta-community landscape (Gravel et al. 2011; Trøjelsgaard & Olesen 2016), and their structure both locally and globally, is constrained by, but is also a constraint on, co-occurrence (Araújo et al. 2011; Cazelles et al. 2015).

Analyzing networks in a meta-community context might require a new representation. Most of this challenge comes from two sources. First, species are shared across locations; this means that two nodes in two networks may actually represent the same species. Second, networks are connected by species movement. Both the dynamics and the structure of networks are impacted by the fact that species move across the landscape at different rates and in different ways. The implication is that every species in the landscape potentially experiences its own version of the metacommunity (Olesen et al. 2010). Investigating community structure and the emerging dynamic processes in light of space would allow a more potent examination of the spatial structure and dynamics of ecological networks (Gouhier et al. 2010; Gravel et al. 2016b; Trøjelsgaard & Olesen 2016).

3.4   Managing species interactions networks data

The above analyses benefit from access to (context-enhanced) data on ecological interactions. An important point to raise is that the format expected for the analysis (*i.e.*, when data are actively being processed) is different from the format suitable for storage, archival, mining, and

linking. From an information management perspective, this places the question of What are ecological networks? in a new light.

Most of the measures mentioned above, and therefore most software, expect networks to be represented as matrices; every row/column of the matrix is an object, and the value at row i and column j is a measure of the interaction between i and j. It can be a Boolean value, a measure of interaction strength, or a probability of interaction. This approach is used by databases such as IWDB, Web-of-Life.es, and World of Webs (Thompson et al. 2012). Although this approach has the benefit of being immediately useful, it lacks the easy homogeneousaddition of metadata. In the context of species interaction networks, metadata is required at several levels: nodes (species, individuals), interactions, but also the overall network itself (date of collection, site environmental data, and so on). Most research has so far been constrained to the adjacency matrix representation of networks. However, ontologically richer representations (graphs with built-in metadata) may offer themselves to a larger and different tool set: multi-graphs, and hyper-graphs, capture a wider picture of ecosystems where all types of interactions are considered simultaneously. Food webs, or other networks, stored as binary or weighted matrices may not be the most relevant representation for these questions.

There are two initiatives that remedy this shortcoming by providing meta-data-rich informa-tion on ecological interactions. `Globi` (Poelen et al. 2014) is a database of interactions, extracted from the literature, and available through GBIF. It relies on an ontology of interaction types, and on unique taxonomic identifiers for species. `Mangal.io` (Poisot et al. 2015a) is a database of networks, that can be fed and queried openly through several packages; it relies on a custom data format, and can be linked to other databases through the use of taxonomic identifiers.

Networks formatted as raw matrices may well be immediately usable, but supplementing them with external information is hard. On the other hand, granular databases with rich metadata can always be converted to raw matrices, while retaining additional information. It

84

is important that we maintain a distinction between the formats used for storage (in which case, relational databases are the clear winner) and the formats used for analysis (that can be generated from queries of databases). In order to facilitate synthesis, and draw on existing data sources, it seems important that the practice of depositing interaction matrices be retired, with the profit of contributing to the growth of context-rich databases. here are a handful of software packages available for ecological network analysis (Csardi & Nepusz 2006; Dormann et al. 2008; Hagberg et al. 2008; Hudson et al. 2013; Flores et al. 2016; Poisot et al. 2016a). They differ in their language of implementation, license, and methods availability.

Considerations about the analysis of networks go hand in hand with the far more difficult question of data sources and data quality. Jordano (2016) showed that obtaining estimates of the completeness of sampling is both difficult, and different between weighted and unweighted networks. Describing the data at the level of the interaction in more detail may therefore give better estimates of (i) the robustness of the overall network, and (ii) the relevant aspects of life history to add in models. These can then be added to predictive models, in the form of functional traits (Bartomeus 2013; Bartomeus et al. 2016), to boost our ability to infer the existence of interactions (or their strength). Relevant interaction-level data (discussed in Poisot et al. 2015a) include the identity of species involved, their abundances, local environmental conditions, and functional traits of the individuals or populations observed interacting, when available. Shifting the focus of sampling away from networks, and onto interactions (because what are networks, but a collection of interactions?) would give more information to work with. Because the amount, resolution, and type of information that it is necessary and feasible to sample will vary for each system, empirical network scientists should lead the effort involved with developing data standards. Taking a step back, data quality should be framed within the context of a specific analysis; we feel that there is a need for a review that would attempt to determine the minimal amount of information needed as a function of the type of analyses that will be applied.

3.5 Conclusions

In this contribution, we have attempted a summary of the measures from graph theory that are the most frequently used in, or the most relevant to, the analysis of species interaction networks.

Even though species interaction networks are ubiquitous in community ecology, biogeography, macroecology, and so on, there is no clear consensus on how to analyse them. We identified a number of areas that would benefit from methodological development. We highlight each of these below, and identify whether they should stimulate future development of novel methods to complete the framework, or stimulate further investigation and assessment of existing methods to clarify when they should be applied.

There is a pressing need to accommodate hypergraphs and multigraphs within the network analysis framework, to allow work on a larger and more realistic variety of ecological situations. Pilosof et al. (2015) identified these systems as having a high relevance when predicting community change, and the emergence of zoonotic diseases, and this is a clear example of an area in which ecology and applied mathematics can have fruitful interaction.

The information we use in the building of network needs to be expanded. Far from being a collection of species and their interactions, networks are structured by environmental forces, species trait distribution, species evolutionary history, and random chance. Replicated data sets with extensive metadata and additional information would most likely boost our power to describe, explain, and predict network structure (Poisot et al. 2016b). The next generation of network measures should account for additional information carried by both species and interactions.

Of course, the addition of data to ecological interactions requires to expand the scope of what is currently being sampled, and to normalize it to some extent. More broadly, we expect that the development of novel methods, and the collection of novel data and their standardization, should go hand in hand. The emergence of interactions and networks databases,

based around documented formats, is a step in the right direction, as they provide an idea of the scope of data to collect.

We need to establish stronger standards for the manipulation of network data. Networks are difficult to manipulate, and the lack of a robust software suite to analyse them is a very worrying trend – our knowledge of ecological networks is only as good as our implementation of the analyses, and academic code can always be made more robust, especially in fields where the widespread adoption of computational approaches is still ongoing. We expect that, since there are numerous initiatives to increase good practices in software engineering among academics, this problem will be solved by improved community standards in the coming years.

The NHST approach to network structure needs additional study, especially when it comes to determining best practice. Recent developments in graph theory, and notably edge-sampling based cross-validation (Li et al. 2020), can help assess the performance of generative null models. There is a shortage of null models that are based on topology but still account for known biology of the networks (such as forbidden interactions), highlighting the need for future developments.

There is a need to compare the alternative measures of a single property. We tried as far as possible to frame these measures in the context of their ecological meaning. But this can only be done properly by strengthening the ties between network analysis and field- or laboratory-based community ecology. Statistical analysis of measures on existing data sets will only go so far, and we call for the next generation of studies aiming to understand the properties of network structure to be built around collaboration between empirical researchers and measures developers.

Chapitre 4

Bactéries et phages en milieu contrôlé

4.1   Coévolution expérimentale

Ce chapitre est un peu particulier puisqu'il revient sur un pan de mon doctorat qui a été abandonné. Cependant, il m'a permis d'avoir une meilleure compréhension du système bactéries-phages, ce qui me servira par la suite pour l'élaboration du modèle utilisé dans le [chapitre 5][Bacteria and Phage coevolution: Phage generality impact on network strucutre]. À l'origine, ce chapitre s'insérait dans un projet qui visait à identifier l'effet de la température sur la coévolution entre les bactéries et les phages, et ce de façon expérimentale et théorique. Plus particulièrement, ce projet portait sur les capacités adaptatives et coévolutives des bactéries et des phages en laboratoire, en fonction de la température de l'environnement. Il était question de voir qu'il été possible de noter une accélération de la coévolution avec la température, mais également la vitesse d'adaptation locale des deux organismes. Le choix du modèle biologique était stratégique car, comme nous allons le voir par la suite, ce sont des organismes qu'il est possible de faire grandir en laboratoire et dont les générations permettent de voir l'évolution en temps réel. C'est également un modèle qui a fait l'objet d'études en écologie et évolution auxquelles nous aurions donc pu confronter nos résultats. Je vais vous présenter ici les démarches qui ont été entreprises et le cul-de-sac dans lequel je me suis retrouvée.

4.1.1   Les bactéries et phages de la plante Sarracenia purpurea

Dans le domaine de l'écologie évolutive, de nombreuses études portant sur la coévolution des bactéries et des phages utilisent le couple *Pseudomonas fluorescens* et le phage associé $\phi 2$ (Buckling & Rainey 2002 ; Brockhurst et al. 2003a ; Hall et al. 2011c ; Poisot et al. 2011 ; Scanlan et al. 2015 , 2016; Gurney et al. 2017). Cependant, du fait que je voulais explorer l'effet de la température dans le processus d'adaptation locale et de coévolution, j'ai utilisé des bactéries et des phages de la plante insectivore *Sarracenia purpurea* (L.) (fig. 4.1).
S. purpurea a en effet une très large distribution géographique, s'étendant du nord du Québec à la Floride. Cette distribution offre un large gradient naturel de température et fait de la

sarracénie pourpre un choix idéal pour mener mes expériences. Je pouvais alors combiner des bactéries et des phages provenant de différentes localités sur le gradient et noter s'il y avait une différence dans leur capacité d'adaptation.

Cette plante est facilement reconnaissable à sa feuille en forme de tube. Pendant la saison estivale, les tubes formés par les nouvelles feuilles s'ouvrent et se remplissent d'eau de pluie et de ses micro-organismes. Les communautés bactériennes contenues dans le tube de *S. purpurea* sont d'une importance cruciale pour le maintien de la population végétale. En effet, les sols des zones humides, où poussent *S. purpurea*, ont un faible taux de nutriments. Afin d'assurer un niveau d'azote et de phosphore suffisant pour leur croissance et son développement, *S. purpurea* a développé des feuilles en forme de tube où les insectes se retrouvent piégés (Whitman et al. 2005). L'eau de pluie recueillie dans ces "tubes" ainsi qu'un sucre sécrété par la plante en font un piège passif efficace dans lequel se noient les insectes qui s'y sont aventurés. Ceux-ci sont ensuite déchiquetés par des mouches saprophages. Leurs débris sont alors interceptés par des bactéries qui, en les dégradant, libèrent l'azote nécessaire au développement de la plante (Butler & Ellison 2007). Cet azote est, par la suite, récupéré par la plante via les parois de la cavité foliaire (Prankevicius & Cameron 1991). La communauté bactérienne présente dans l'eau de *S. purpurea* est régulée par le réseau trophique qu'on y trouve, via un effet top-down, notamment avec l'influence de la larve de moustique *Wyeomiia smithi* (Kneitel & Miller 2002) ainsi que des rotifères et protozoaires (Baiser et al. 2011).

Cette chaine alimentaire et ce qu'il apporte à la plante ont été largement étudiés (Cochran-Stafira & von Ende 1998 ; Kneitel & Miller 2002 ; Baiser et al. 2011; Friman & Buckling 2013). Cependant, le rôle des phages dans la régulation des bactéries est encore mal connu. L'utilisation des phages et des bactéries de cette plante permettrait alors d'accroître les connaissances que nous avons déjà sur cet incroyable écosystème.

Figure 4.1 Photo d'une Sarracenia purpurea prise durant l'échantillonnage au Québec en juillet 2016. Plusieurs feuilles tubuleuses sont visibles au pieds des fleurs, ces feuilles compose ce que l'on appelle une rosette.

4.1.2 Échantillonnage

L'échantillonnage a été réalisé en collaboration avec plusieurs laboratoires (Baiser Lab of Community Ecology - Benjamin Baiser, University of Florida ; Dominique Gravel Laboratory, Université de Sherbrooke ; Zac Freedman Laboratory, University of Michigan ; The Record Lab - Sydne Record, Computational Ecology at Grym Mawr College). Ces laboratoires sont liés par un projet commun visant l'accroissement des connaissances sur *S. purpurea*. Un protocole commun nous permet de réduire l'effort d'échantillonnage et de couvrir une large zone d'étude en peu de temps. Les sites d'échantillonnage ont été choisis, sur la base d'études précédentes, lors d'un atelier en janvier 2016. Ceux du Québec ont été choisis à l'aide de la base de données [GBIF][https://www.gbif.org/] et pendant le temps d'exploration de l'échantillonnage, en juillet 2016.

Deux équipes différentes ont échantillonné des données : l'une aux États-Unis, l'autre au Québec. Une session de partage des techniques a été faite pour standardiser au maximum l'échantillonnage. Grâce à cette collaboration, les échantillons récoltés ont permis de mener une étude sur la distribution des bactéries présentent dans cet écosystème (*Annexe B*).

Au final, 36 sites ont été échantillonnés, répartis entre le nord de la Floride et la Baie-James au Québec (fig. 4.2). Sur chacun d'eux, un transect d'au moins 120 mètres a été tracé au travers des populations de *S. purpurea*. Tous les 5 mètres, dans la rosette (*i.e.* bouquet de feuilles) la plus proche du transect, nous avons échantillonné la plus grande des feuilles nouvellement ouvertes. L'âge de la feuille a été déterminé par son épaisseur et sa rigidité. Un maximum de 24 feuilles a été échantillonné, lorsque la population était suffisamment grande. Toutes les 5 plantes, l'eau de la feuille-tube a été collectée puis filtrée pour obtenir 1,6 ml de solution de micro-organismes. Lorsque les sites étaient trop secs, la récolte de l'eau se faisait dans toute feuille pleine d'eau, quelle que soit sa position sur le transect. Dans chaque solution filtrée, nous avons ajouté du glycérol à 67% et nous les avons placés à $-20°C$ pour le reste de la session d'échantillonnage et à $-80°C$ de retour au laboratoire.
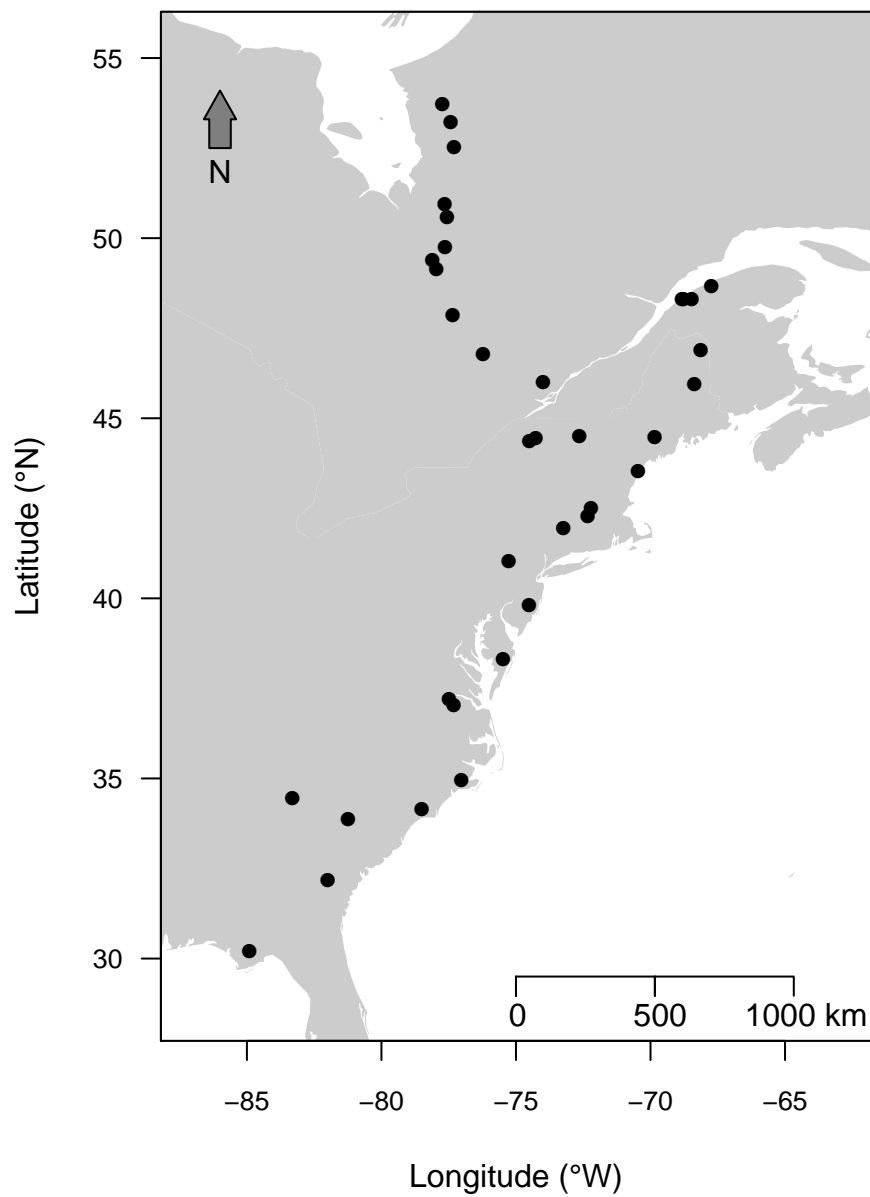
Figure 4.2 Distribution spatiale des sites d'échantillonnage. Du nord de la Floride (USA) à la Baie-James (Québec, Canada). Chaque point représente un site d'échantillonnage (n = 36).

4.1.3 Expérimentations en laboratoire

En me basant sur les méthodes décrites par Buckling & Rainey (2002) et Poullain et al. (2008), je voulais alors cultiver des populations de bactéries et de phages dans différentes conditions de température et tester leur résistance et leur infectivité, mais c'est ici que les choses se sont gâtées.

L'isolement des bactéries a été un succès, cependant, celui des phages s'est avéré un peu plus compliqué. Avant de revenir sur les obstacles rencontrés durant l'isolation des phages, il convient de détailler le protocole expérimental qui avait été mis en place pour tester le fonctionnement de l'isolation de phages et des bactéries.

Protocole initial   Le protocole d'isolement des phages et des bactéries a été construit à partir des informations recueillies dans la littérature. L'approche présentée ici a servi de test pour évaluer la faisabilité du protocole général qui sera utilisé pour les expériences de coévolution jusqu'à lors prévu. Les tests d'isolations des bactéries et des phages sont été divisées en cinq étapes principales : (1) amplification de l'échantillon, (2) centrifugation pour séparer les phages et les bactéries, en utilisant ou non du chloroforme à 10, (3) isolement et amplification du type morphologique bactérien, (4) dilution en série des phages et (5) mesure de la croissance bactérienne (avec la densité optique) avec et sans phages.

Pour des raisons de simplicité, j'ai testé mon protocole avec 3 échantillons provenant (i) de latitudes nord, (ii) de latitudes moyennes et (iii) de latitudes sud. J'ai choisi de ne pas utiliser le chloroforme pour séparer les phages et les bactéries, à la place j'ai réalisé deux centrifugations consécutives. Ce protocole initial est représenté dans la figure 4.3.

Isolation et amplification des bactéries   L'isolement et l'amplification des bactéries ont été réalisés relativement facilement. Les types morphologiques bactériens se développant dans les conditions de laboratoire étaient généralement aux environs de deux. Au début, les deux types
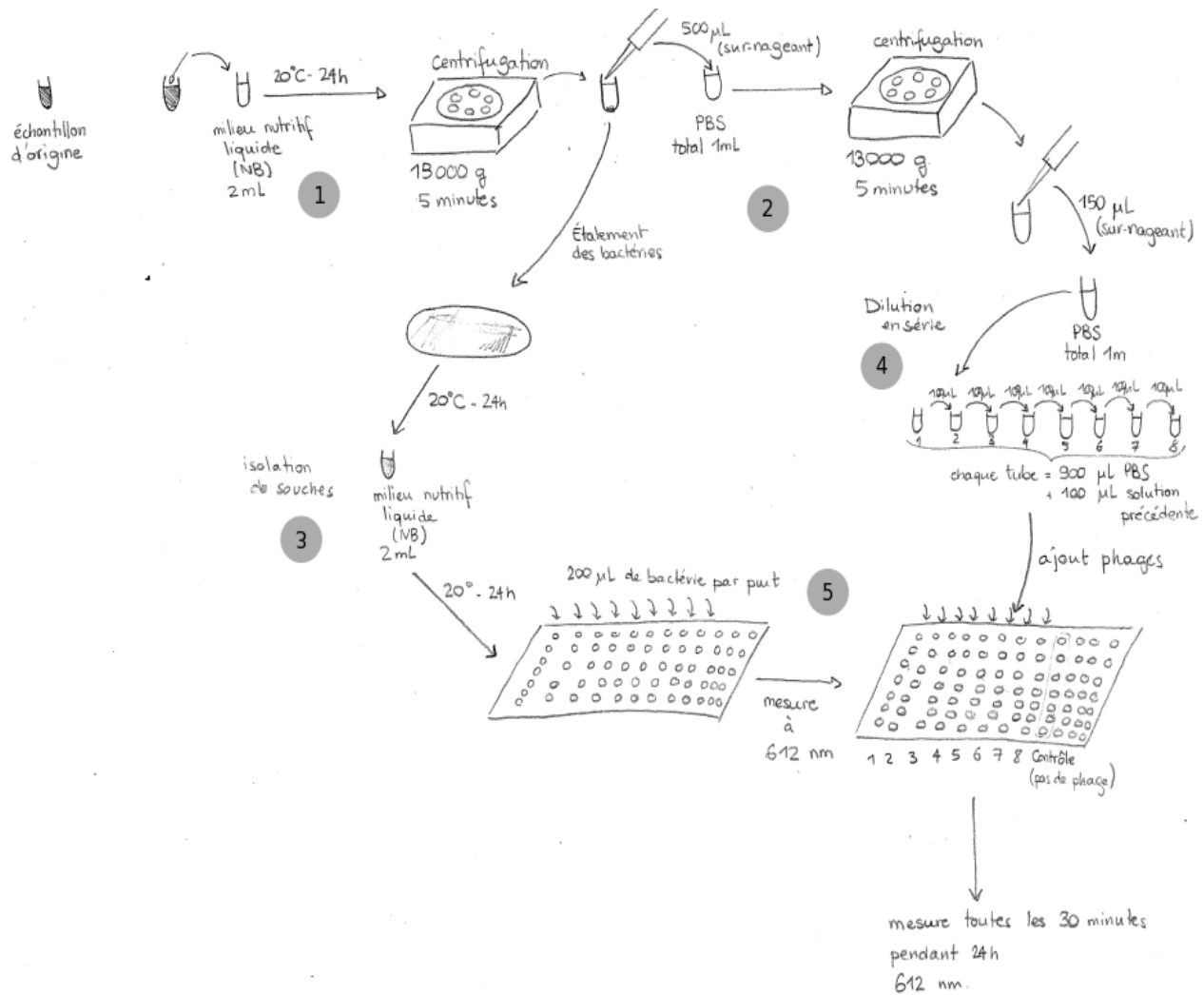
Figure 4.3 Protocole expérimental pour l'isolation des phages et des bactéries présentes dans les échantillons. Ce protocole est divisé en cinq étapes : (1) amplification des échantillons, (2) centrifugation pour séparer les phages des bactéries, avec usage ou non de chloroforme 10, (3) isolation et amplification des morphotypes de bactéries, (4) dilution en série pour les phages jusqu'à une dilution à 10^{-8} et (5) mesure de la croissance bactérienne avec et sans phages, grâce à l'utilisation d'un spectromètre à 612nm toutes les 30 minutes pendant 24 heures, le tout continuellement agiter pour favoriser les interactions entres phages et bactéries.

morphologiques ont été testés, puis un seul des types morphologiques bactériens actuels a été utilisé pour réaliser les tests. Les types morphologiques bactériens ont été choisis au hasard, à pile ou face.

Isolation des phages et test d'infectivité   Après la première exécution du protocole, les mesures spectrophotométriques (à 612 nm) de liquide de culture, dans lequel se trouvait supposément des bactéries et des phages, n'ont pas permis de révélé une quelqu'onque action de ces derniers sur le développement bactérien. Aucune interaction phage-bactérie n'a été mise en évidence. Ce résultat peut être causé par deux facteurs : (i) la présence de bactéries dans la solution phagique, diminuant drastiquement l'effet des phages sur les bactéries isolées ou (ii) l'absence de phages dans la solution phagique.

Y a-t-il des bactéries dans les solutions phagiques ?   Pour tester si des bactéries étaient encore présentes dans la solution de phage, j'ai refait les étapes 1 et 2 du protocole initial. Puis j'ai ensemencé sur un milieu nutritif (formé avec du NB, Nutrient Broth) la solution de phage obtenue après deux centrifugations consécutives de 5 minutes à 13000g. Après 24 heures à $20°C$, les tests étaient positifs à la présence de bactéries.

- Tests de séparation

Pour trouver un moyen d'éliminer les bactéries dans la solution de phage, plusieurs tests ont été réalisés. Pour chaque test, les étapes 1 et 2 du protocole initial ont été répétées.

À la lumière de la littérature sur la séparation bactéries-phages, j'ai décidé de tester deux forces de centrifugation (13000g et 15000g), avec des temps de centrifugation allant de 10 à 20 minutes et des temps d'incubation de 24 ou 48h. Le volume du surnageant collecté a également été testé (premier prélèvement : 750, 500, 250, 150 μl). Malheureusement, aucun de ces tests n'a permis de mettre en évidence la présence de phage dans les solutions.

- Ajout d'un agent bactéricide

1. L'éthanol

Utilisé comme antibactérien dans de nombreux endroits, que ce soit dans les hôpitaux, les laboratoires ou d'autres lieux publics, l'éthanol peut détruire une grande partie des bactéries présentes dans un milieu. Il agit en dénaturant les protéines membranaires des bactéries et en dissolvant les lipides présents dans cette membrane (Ingram & Vreeland 1980). L'éthanol est le plus efficace à 70% (Malik et al. 2006), j'ai donc commencé les tests en ajoutant 10% au surnageant entre les deux centrifugations. Après la deuxième centrifugation, le surnageant est étalé sur un milieu nutritif pendant 24 à 48 heures à 20°C.
Malheureusement, les tests incluant l'éthanol n'ont pas permis d'obtenir une preuve d'isolation réussie de phages.

2. Le chloroforme

Le chloroforme est couramment utilisé pour éliminer les bactéries indésirables des solutions de phages. Il détruit les membranes cellulaires des bactéries, ce qui les conduit à une mort certaine. J'ai donc commencé les tests en ajoutant 10% de chloroforme au surnageant prélevé entre les deux centrifugations. Après la deuxième centrifugation, le surnageant est étalé sur un milieu nutritif pendant 24 à 48 heures à 20°C.
Cette fois, les solutions de phages obtenues avec l'ajout de chloroforme à 10 ont été retrouvées sans aucune trace de bactéries vivantes. Le chloroforme a donc été ajouté au surnageant avant la deuxième centrifugation dans les étapes du protocole, ce qui permettait de préserver, d'isoler et d'amplifier les phages pour la suite de l'expérience.

Y a-t-il des phages dans les solutions phagiques ? Maintenant que la question des bactéries dans la solution de phages est réglée, il paraissait important de savoir si des phages avaient pu être isolés.

- Spot tests

Après avoir amplifié l'échantillon original, puis séparé les phages et les bactéries par centrifugation, j'ai réalisé des tests ponctuels. Il s'agit de préparer un milieu de culture de bactéries inoculées sur lequel on applique des gouttes de solution phagique à différentes concentrations. Après une incubation à $20°C$ pendant 24 à 48 heures, les concentrations pour lesquelles les phages sont capables d'interagir avec les bactéries devraient être visibles grâce à l'inhibition de la croissance bactérienne.

Après plusieurs essais sur la méthodologie utilisée pour réaliser les tests ponctuels, seule la concentration initiale de la solution de phages a eu un effet inhibiteur sur la croissance bactérienne.

Ce résultat indique soit (i) que le chloroforme est en concentration suffisante pour avoir détruit les bactéries présentes dans le milieu et/ou (ii) que la quantité de phages présente dans les solutions phagiques est trop faible pour avoir un impact sur la croissance des bactéries, quelle que soit la concentration de la solution phagique.

- La microscopie à fluorescence

Afin de détecter la présence de phages dans une solution, il est possible d'utiliser la microscopie à fluorescence. Le colorant SYBR-Green se lie aux acides nucléiques et devient fluorescent. Son utilisation permet la détection de l'ADN présent dans les capsides des phages. La fluorescence émise a une durée de vie limitée, mais elle permet tout de même de voir s'il y a ou non des phages en suspension dans la solution. Grâce à l'aide du coordinateur des machines du département de biologie de l'Université de Sherbrooke, j'ai pu inspecter deux échantillons de solution de phages. Aucun phage n'y a été détecté.

4.1.4   Quelles conclusions en tirer ?

Les échantillons qui ont été utilisés ont un faible volume (1,6 ml prélevé par plante, sur lequel nous prenons 4 μl pour faire la culture et l'amplification en laboratoire) contrairement aux échantillons utilisés pour l'isolement des phages dans la littérature. Cette différence de volume

peut expliquer la raison pour laquelle je n'ai pas réussi à détecter de phage dans mes solutions phagiques. Cependant, dans le but d'y voir un peu plus clair, j'ai pris la décision d'en discuter avec un professeur en virologie de l'Université de Sherbrooke. Après avoir expliqué ma situation, celui-ci me répond simplement qu'au vu du système avec lequel je travaille, à savoir le système interne au tube des sarracénies, qu'il est normal de ne pas trouver de phages dans mes solutions. Il s'explique : le système des sarracénies est un système clos, avec un apport discontinu en ressources, que ce soit insectes ou bactéries. Pour cette raison, les phages qui pourraientt potentiellement subsister dans ce genre de milieu sont des phages lysogéniques. Or, le type de protocole expérimental utilisé pour étudier la coévolution entre les bactéries et les phages requiert des phages lytiques.

Après la réception de cette information, qui ne m'était pas passée par l'esprit, j'ai décidé de ne pas persévérer dans la partie expérimentale de mon doctorat. Bien qu'il ne soit rien ressorti de ce projet, il m'a permis de m'imprégner des cycles de vie des bactéries et des phages et de la façon dont les expérimentations en laboratoire peuvent se dérouler. Grâce à cela, j'ai pu terminer de construire un modèle mathématique recréant virtuellement les études expérimentales de coévolution entre phages et bactérie, lequel est utilisé pour le chapitre suivant.

Chapitre 5

Bacteria and Phage coevolution: Phage generality impact on network structure

---

Bacteria and Phage coevolution: Phage generality impact on network structure

Mathilde Besson[1,2], Dominique Gravel[2,3] & Timothée Poisot[1,2]

[1] Département de Sciences Biologiques, Université de Montréal, Montréal, H2V 2J7, Canada; [2] Québec Centre for Biodiversity Sciences, McGill University, Montréal, H3A 1B1, Canada; [3] Département de Biologie, Université de Sherbrooke, Sherbrooke, J1K 2R1, Canada

in prep.

Abstract

Bacteria-phage systems are common biological models for the study of coevolutionary dynamics. To quantify and interpret their evolution in a controlled environment, time-shift experiments analysis are widely used, because it allows us to determine the adaptive dynamics between a parasite and its host. Beyond the changes in infection and resistance at the community level, the finer structure of the interaction networks between bacteria and phages is also a key element in their coevolution. This structure is indeed correlated with the evolutionary dynamic involved. In this paper, we examine whether network structure metrics can be good indicators of the coevolutionary dynamics between bacteria and phages. We further evaluate the impact of certain latent phage traits on evolutionary dynamics.

Using an individual based model, we reproduced the dynamics of growth, competition, and infection between bacteria and phages. We extracted information on (i) the duration of simulations, (ii) the trait distribution of individuals over time, and (iii) their interactions. We investigate the variation of these three indicators as a function of traits phages, such as virulence and generality. We then performed time-shift experiments using individuals traits and analyzed the coevolutionary dynamics coming out from the simulations. We also calculated the nestedness over time to evaluate if network structure can be a representation of the coevolutionary dynamics.

The simulations results show that the rate of coevolutionary dynamic between phage and bacteria decreases over time, accompanied by a decrease in nestedness. Phage generality appears to be a key factor determining coevolutionary rate and interaction network structure: the more generalist the phages are, the higher the coevolution rate and the nestedness. We could not see the same clear tendency in the influence of phage virulence. Virtual time-shift experiments indicate that the bacteria-phage system is guided by an arms race dynamic. Conversely, the evolution of nestedness would tend to show a fluctuating selection dynamic. We conclude that virulence is not necessarily the most important key trait in the evolutionary

dynamic between bacteria and phage, phage generality seems to be the one. Also, in the determination of evolutionary dynamics, nestedness is not necessarily suitable as a measure of evolutionary dynamics between bacteria and phages considering the mismatching of the dynamic found with time-shift experiments.

5.1 Introduction

For several decades now, bacteria and phages have been, for good reasons, model organisms in the study of evolutionary dynamics. In particular, they are extremely useful to link biological processes to community structure (Bohannan & Lenski 2000), but also as biological models to understand the processes underlying coevolution (Buckling & Brockhurst 2012). Their strength lies in the fact that they are very easily grown in the laboratory and have a relatively short generation time that allows us to follow in real time their evolutionary dynamics in place in a community. They can also be stored for later comparison with their ancestors and descendants. It is thanks to this specificity that cross-infection experiments, in time or space, were born (see Koskella 2014) and have been since then largely studied and validated (Gandon et al. 2008). The time-shift experiments highlighted the different dynamics at work in time between bacteria and phages. It is therefore now possible to easily visualize coevolutionary dynamics. For instance, Hall et al. (2011d) found that coevolution between phages and bacteria shows different dynamics over time, starting with an arm race dynamic (ARD) and then quickly shifting to a fluctuating selection dynamic (FSD). The environment also tends to be a determinant factor in the dynamic that can be observed (Brockhurst et al. 2003b). For instance, coevolution in controlled conditions (laboratory) shows most of the time an ARD (Brockhurst et al. 2003a), while in natural conditions, like in soil, it shows FSD (Gómez & Buckling 2011). The combination of time-shift experiments and the laboratory protocol started with Buckling and Rainey (2002) and then served as a prolific cocktail for the accumulation of knowledge about coevolutionary dynamics. For example, Poullain et al. (2008) showed that the

rate of phage evolution differed depending on whether the bacteria were allowed to evolve or not, which, among other things, influenced phage growth and infectivity. Bacteria and phages being relatively simple to model, different mathematical models have emerged, fleshing out theories about evolutionary dynamics. Using an explicit trait-driven model, Weitz et al. (2005) tested the possible coexistence and coevolution of phage and bacteria in a laboratory context. From now, phage and bacteria seem to be a well-known biological and mathematical model, but laboratory experiments and mathematical models are usually made in different ways and purposes. Experiments could increase largely their understanding of what exactly is happening during evolutionary processes using models.

One of the inescapable steps of coevolutionary studies lies in these analyses and indicators used. As we saw just before, the cross-infection experiment has been widely used for this purpose. The links between coevolution and network structure emerged and have been investigated for several decades now. It is also possible to observe the two different types of network structures, nested (*i.e.* generalists species interacting with more specialists, and so on) and modular (*i.e.* group of species with more interactions than expected and than with the rest of a larger group), generally associated with divergent situations. It is now known that interactions influence coevolution between bacteria and phages (Andreazzi et al. 2017), and that the type of structure will have a different impact on coevolution and vice versa. For example, coevolutionary processes enhance changes in trait matching, resulting in the evolution of interacting species, which in this way will influence the possible coevolution. However, the link between the two is still unclear, especially when we talk about laboratory experiments, since the culture conditions are far from approaching natural environments, and the fact that measuring coevolutionary dynamics is still a challenge (Hall et al. 2020). Studies on the relationship between the structure of interaction networks are still too few, especially experimental studies (Fortuna et al. 2019), and above all show, sometimes, contradictory results. For example, Fortuna et al. (2019) find that ARD is associated with increasing nestedness over time, and vice versa for FSD.

Only a few studies focused on this particular topic, so it would be interesting the evaluate the relevance of the network structure as an indicator of coevolutionary dynamics. In this quest for clarification, theoretical studies can help, providing insight into what happens under the surface. But the diversity of models available can sometimes be more blurring than clarifying.

When we focus on other characteristics that could influence coevolution between phages and bacteria, virulence and generality of phages are two traits that stand out as having a significant influence. Indeed, Bull (1994) shows that the virulence of the parasite is a clever trade-off between the negative impacts it can have on the host and the benefits for the parasite: too much virulence drastically increases mortality in the hosts, limiting parasites dispersal. However, in a coevolutionary context, in experiments where two populations of parasites and their host grow together, there is a tendency to see that parasites become more and more virulent over time as host resistance increases (Thrall & Burdon 2003). Phage generality also plays a major role in the coevolutionary processes as the generality of phages increases in time when they can evolve with bacteria (Hall et al. 2011a). But this phage trait is conditioned by external conditions as the level of resources impacts the level of phage generality (Poisot et al. 2011). Moreover, there is a trade-off between the virulence of phage and their generality, highlighted by cross infection experiments. When phages are generalist, they tend to be less virulent, which allows them to infect more hosts. Knowing that these traits, virulence, and generality, can both influence coevolution in their way, it would be interesting to track the effect of each on coevolutionary dynamic metrics, which would be useful when planning an experiment.

From what we know about the network structure and properties of bacteria and phages, we can expect that the structure of the networks varies according to the different characteristics of the phages. Moreover, if this structure evolves over time, we should be able to draw inferences about the evolutionary dynamics between bacteria and phages. For this purpose, we used a mathematical model that allows us to follow the evolution and the possible coevolution of

bacteria and phages. As studies taking into account the structure of interaction networks as an indicator of evolutionary dynamics are still scarce, we used data from time-shift cross-infection experiments, which are known to have been successful, to validate the results obtained with the structure of networks. Thanks to these two indicators, time-shift experiments, and network structure, we will finally be able to observe the effect of virulence and generality of phages on coevolutionary dynamics with bacteria.

5.2   Methods

To understand what's going on during bacteria-phage coevolution experimentations, we built a mathematical model which reproduces bacteria and phage growth in a controlled environment and allows us to follow their evolution and coevolution. This is an individual based model, divided into three different phases: interaction, reproduction, and evolution. We could track evolution across time through individuals traits of phages and bacteria. Thanks to this model, we reproduced time-shift infection experiments and also collected information about bacteria-phage interactions network structure over time, which gives clues about the possible relationship between coevolution and interactions network structure.

5.2.1   Model

Interaction   As we just said, our model is based on bacteria and phages populations where each individual is identified with a trait, represented by a numerical value that we call *trait value*. Traits values help us to organise individuals in a virtual environment and determine if they can interact or not. Interactions can be of two kinds: infection, in case of bacteria-phage interaction, or competition, for bacteria-bacteria interaction. The probability of two individuals interacting is calculated using the absolute distance between individual trait values. The closer they are, the higher the probability of interaction. For bacteria-phage interaction especially, the probability of interaction is different from zero when bacterial trait value enters the phage interaction

range, centered around phage trait value. This probability is called infection probability and is calculated as follows:

$$P(Infection) = exp^{-(D^2)/(2*n^2)},$$ (5.1)

where D is the absolute distance value between phage and bacteria traits and n represents the half of phage interaction range in which it can interact with bacteria. Competition between bacteria is calculated the same way, with an interaction range specific for inter-bacteria interactions.

Individual fitness is calculated from these interactions probabilities. It will be used to create the next generation of bacteria and phages that we'll see in the next phase. Phage fitness is simply calculated with the infection probability (equations 5.3). Bacterial fitness takes into account several other elements (equations 5.2). We added bacterial competition, which can be seen as a competition for resources and then will limit the bacterial growth if the amount of bacteria with very similar trait values becomes too high. It will result in a decrease in reproduction probability of bacteria when the competition is high. We also added an impact of phage virulence on bacterial fitness. The higher the virulence, the lower the bacterial fitness.

$$\omega_{phage} = P(Infection)$$ (5.2)

$$\omega_{bacteria} = 1 - Virulence * P(Infection) - Competition$$ (5.3)

Where Virulence is the phage virulence when they infect a bacterium, and Competition is the intra-bacteria competition.

5.2.2 Reproduction

Individual reproduction is calculated based on fitness and is defined as the number of descendent an individual can produce. Low fitness, for bacteria or phage, will be associated with a low reproduction coefficient and few descendent in the next generation. Phage can produce up to 10 descendants and bacteria only 2, to mimic duplication. Not every individual of this model will reproduce, especially phages. As their replication is dependent on encounters with a compatible bacterium, a non-interacting phage will not be able to reproduce and have descendent in the next generation. The number of descendent is calculated by multiplication of individual fitness and the maximum number of possible descendent. Every created descendent will have the same trait value as its ancestor.

Mutation As we want to have evolution and, hopefully, coevolution in our model, we added a mutation affecting descendent trait values. Each descendant will have a new trait value. This new value is picked in a normal distribution centered around the initial trait value. As phages and bacteria have distinct biological generation times in real life, we applied a different standard deviation of the normal distribution for bacteria and phages (see table 5.1). This allows us to mimic a higher mutation rate for phages due to faster generation time.

5.2.3 Simulations

Simulations were run with Julia (version 1.5.2) and associated code will be available on GitHub (*repository of the PoisotLab*). In order to catch simulations where we could have coevolution between bacteria and phages, only simulations able to run more than 450 time steps are selected for analyses. 450 time steps represent 450 bacterial generations, which is the average amount of generations during Buckling & Rainey (2002) experiments.

A simulation starts with a specific amount of bacteria and phages, N_p and N_b. Each individual (phage and bacteria) receives a trait value. This value is defined randomly from a

normal distribution for each population. The means of these distributions are defined by the parameters $trait0_b$ and $trait0_p$, respectively. The variance is defined at 0.01 for the bacteria population and 0.001 for the phage population. From this point, interaction, reproduction, and mutation phases are iterated at each time step, as explained below.

First, n_t individuals are randomly selected from each bacterial and phage population, where n_t represents the population size of the smaller population at time t. Random couples are made from these sampled population and the probability of infection for each couple is calculated (equation 5.1). The fitness of each individual is then calculated (equations 5.2 and 5.3). Then, reproduction can be enhanced. Each individual trait value is replicated regarding their reproduction and creates a new population, which will be used a time $t + 1$. Finally, individual trait values are slightly changed, using a *mutation effect*, where the standard deviation is defined by the $\sigma_{bacteria}$ and σ_{phage} parameters for bacteria and phages respectively.

Both populations of bacteria and phages are restricted by carrying capacity, respectively K_b and K_p. If the new population outgrows this carrying capacity, individuals are removed at random to reach K_b and/or K_p again. Individual traits are recorded at each time step and the identity of bacteria and phages are selected to interact. With this information, we were able to know how long each simulation lasted and to rebuild interactions networks at any time.

Parameters values that we used are compiled in the table 5.1.

Table 5.1 Model parameters.

Parameter	Signification	Value
N_b	Total number of bacterial individual at the beginning	100
N_p	Total number of phage individual at the beginning	10
K_b	Bacterial carrying capacity	1000
K_p	Phage carrying capacity	5000
$Bdivision$	Number of potential descendant per individual (bacteria)	2
$Pprogeny$	Number of potential virion per individual (phage)	10

$NicheBreadth$	Phage generality	0.05:0.5
$virulence$	Phage virulence	0:1.5
$\sigma_{bacteria}$	Bacterial mutation rate	0.001
σ_{phage}	Phage mutation rate	0.01
$trait0_b$	Initial mean trait value of bacterial population	0.0
$trait0_p$	Initial mean trait value of phage population	0.02

5.2.4   Analysis

At the end of each simulation, we collect information about how long it takes before one of the two populations, bacteria and phages, gets extinct, and which one did extinct first. We also extract information about the' time cross-infection experiment and network structure.

Time-shift cross-infection modeled experiment   For each simulation, we temporarily record every individual traits value for every timestep that the simulation ran. At the end of the run, with this information, we calculate the probability of phage infection on bacteria from several different timesteps. We select phages from t time (75 to 350 with an intervalle of 25 timestep), and calculate the probability of infection on every bacteria from each timestep from t-25 (bacteria from the past) to t+25 (bacteria from the future). This allows us to have a clear image of the evolutionary dynamic happening between our two populations . Then, we keep the mean and standard deviation of population infection probability at the specifics timesteps. The curves obtained with this information give an insight into the evolutionary dynamics involved between the two populations. The slope of these curves indicates the intensity of the dynamic. We extracted the slope using linear regression on the time-shift dynamics results.

Network structure   The nestedness of bacteria and phages networks is used to define the network structure during simulations. As we work with probabilistic bipartite networks, we used

the η function from the package `EcologicalNeworks` (Julia). At every timestep, we randomly selected 100 bacteria and 100 phages and calculated the network nestedness based on their probabilistic interaction matrix. To ensure the nestedness value was representative of the whole network nestedness, we repeated the calculation 10 times and kept the mean as the nestedness value.

5.3   Results

5.3.1   When does coevolution occur?

The simulation time indicates how fast the system between phage and bacteria collapses under a certain set of parameters and under which circumstances we could observe coevolution between phages and bacteria if populations maintain long enough. Figure 5.1 A shows this probability as a function of the generality and virulence of phages. We see that when phages are specialists, no matter how virulent the phage is, the simulation time and the possibility of coevolution are close to zero. As we increase the generality of the phages, we see that the simulations last longer, even reaching an average close to 600 time steps, which is our maximum here, with more generalist phages. However, the more generalist the phages are, the more their virulence will be a non-negligible factor in the simulation time. Indeed, the simulation time decreases drastically as the virulence increases, meaning that the pressure of parasitism is too strong for coevolution to occur. Figure 5.1 B shows that the variation in simulation time is stronger when the virulence pressure of the phage is decisive for the establishment of coevolution. Comparing simulation time to the percentage of phage extinction per parameter set, we can see a clear division of the space. Simulations stopped when one of the two populations gets extinct. In figure 5.1 C, we see that virulent generalist phages tend to get extinct before bacteria.
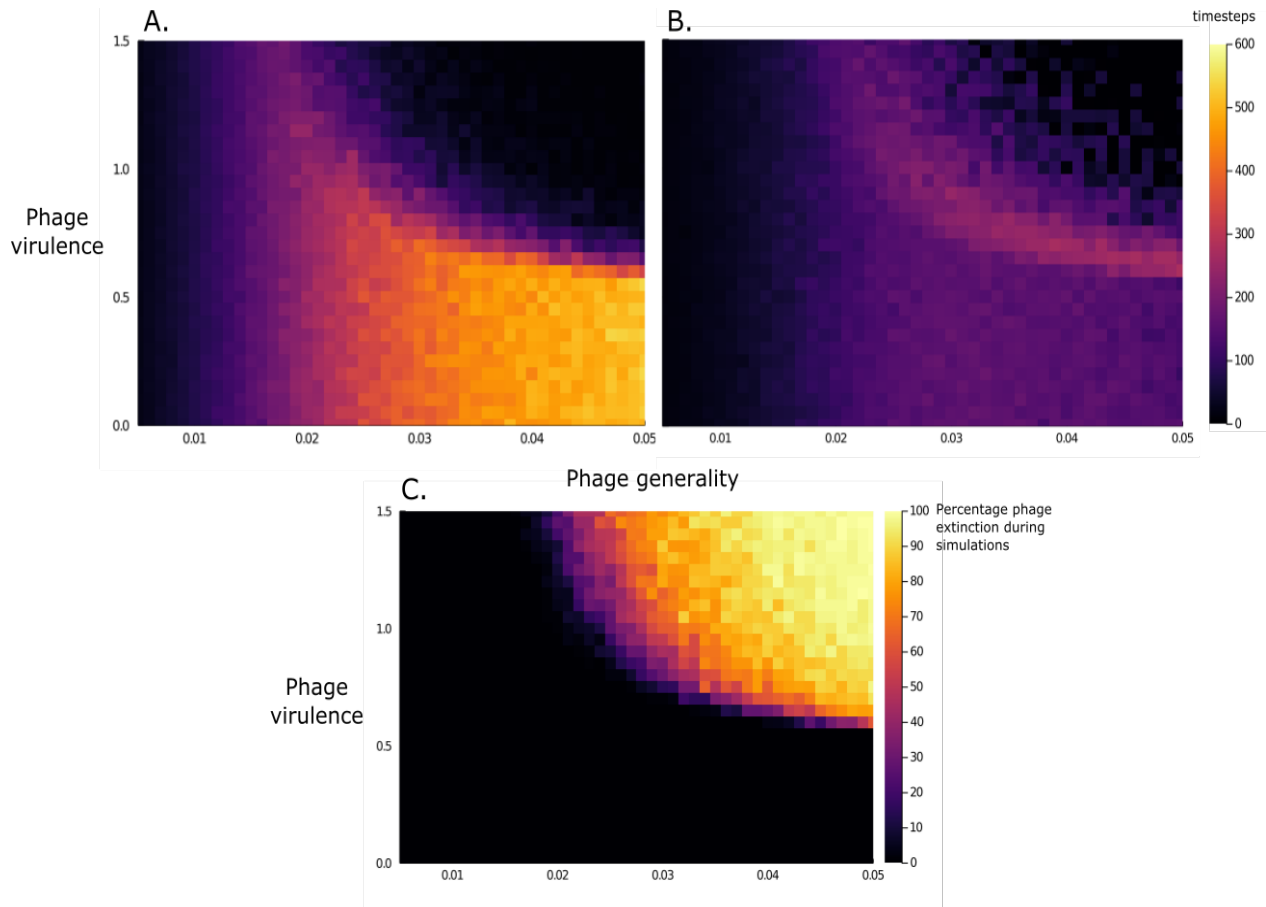
Figure 5.1 Simulations duration depending on phage virulence and generality. A. Simulation mean time. Darker areas represent a combination of phage virulence and generality with simulation runs for only a few time steps, meaning that the bacteria-phage system collapses rapidly. On contrary, brighter areas indicate long-lasting simulations with a maximum represented by yellowish areas. B. Standard deviation of simulation time. Brighter areas represent a wider standard deviation. C. Percentage of simulation ending because of phage extinction.

5.3.2   Which evolutionary dynamic does the time-shift experiment highlight?

The evolutionary dynamic between phages and bacteria is tracked with a time-shift infection experiment. Figure 5.2 shows an example of the dynamic we observe between bacteria and generalist population phages. Here, the probability of infection reflects trait matching between phage and bacteria and indicates the level of adaptation between the two. We see that the mean probability of infection is higher when phages infect contemporary bacteria than bacteria from the past, but this probability is higher for bacteria from the future. This configuration, decreasing linear relationship, is related to the arm race dynamic. We can also see that the mean infection probability decreases over time, but keeps this linear relationship characteristic of the ARD.

To go further and track the evolution of this dynamic depending on time and phage traits, we have extracted the slope of the mean relationship, which was linear, for the different time steps (figure 5.3). On average, the dynamics present negative slopes which represent the arm race dynamic (Gaba & Ebert 2009). For most of the curves, we see that the dynamics slow down with time, as we saw in figure 5.2. The slope of the dynamic is dependent on the phage generality, going from -0.0005 (specialist phages) to -0.001 (generalist phages) at the beginning of simulations (time step 75). For both, this slope becomes close to zero at the end of simulations (time step 350). The evolution of this dynamic is similar regardless of the phage generality. However, it can be noted that the slope increases with the phages generality which implies a stronger coevolutionary dynamic with the presence of generalist phages.

5.3.3   What can network structure tell us?

The structure of phage-bacteria interaction networks also shows changes over time. Even if nestedness is always higher than by random, it changes over time. In figure 5.4 A, we see that it decreases with time, from a mean nestedness of 0.77 ($\pm$ 0.064) at the beginning of the simulations to 0.42 ($\pm$ 0.10) when they have reached structural stability. As for figure
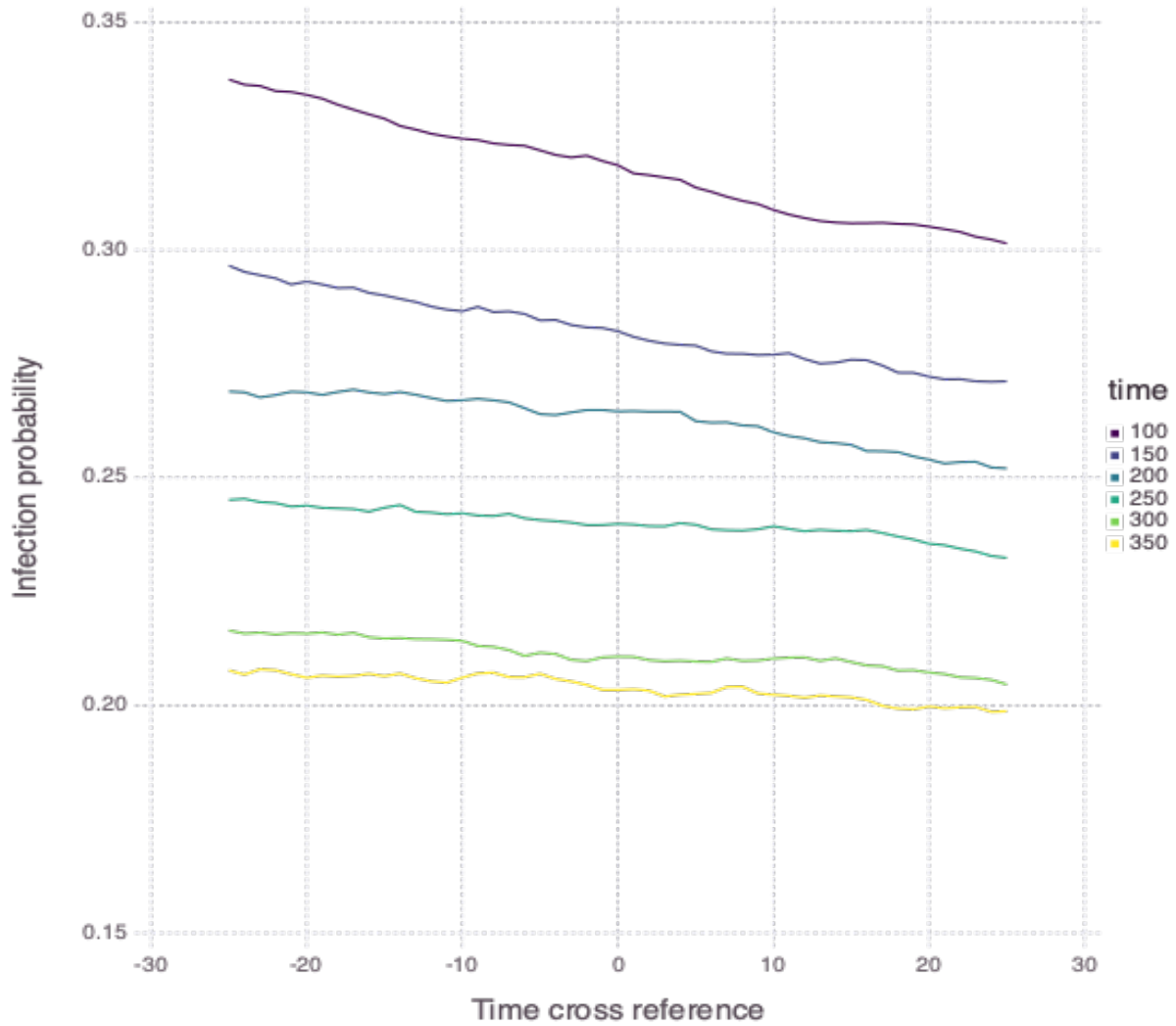
Figure 5.2 Mean infection probability from time-shift cross-infection experiment for generalist phages. On the x axis, 0 représents a couple of contemporary bacteria and phage populations, negative values represent the interaction between the phage population and a bacterial population from the past, and positive values represent the interaction between phage and bacterial population from the future. The colored lines indicate different timesteps during simulation.
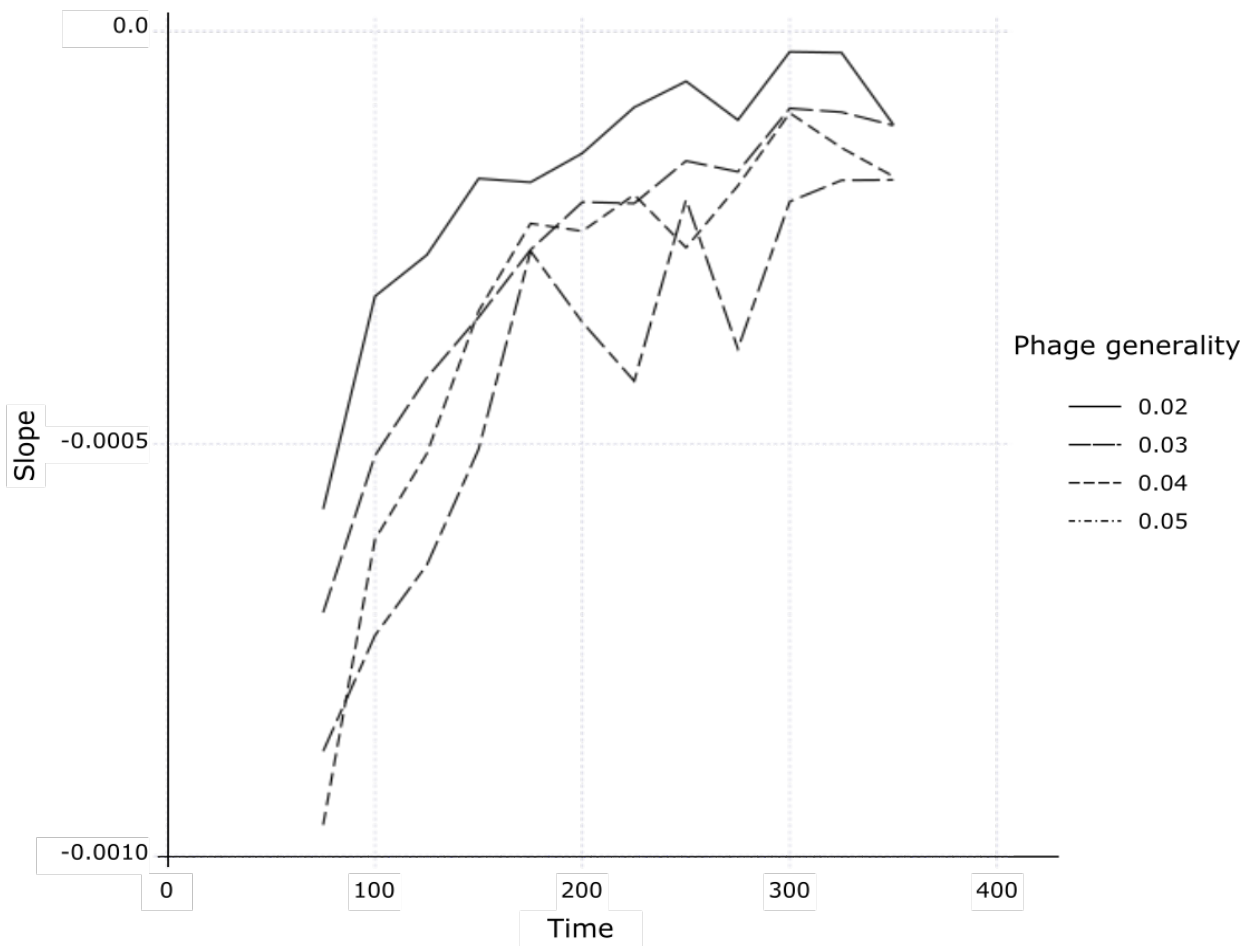
Figure 5.3 Evolution of time-shift cross-infection, over time. Lines styles represent different phage generality. Slope values near zero indicate an evolutionary dynamic slowing down.
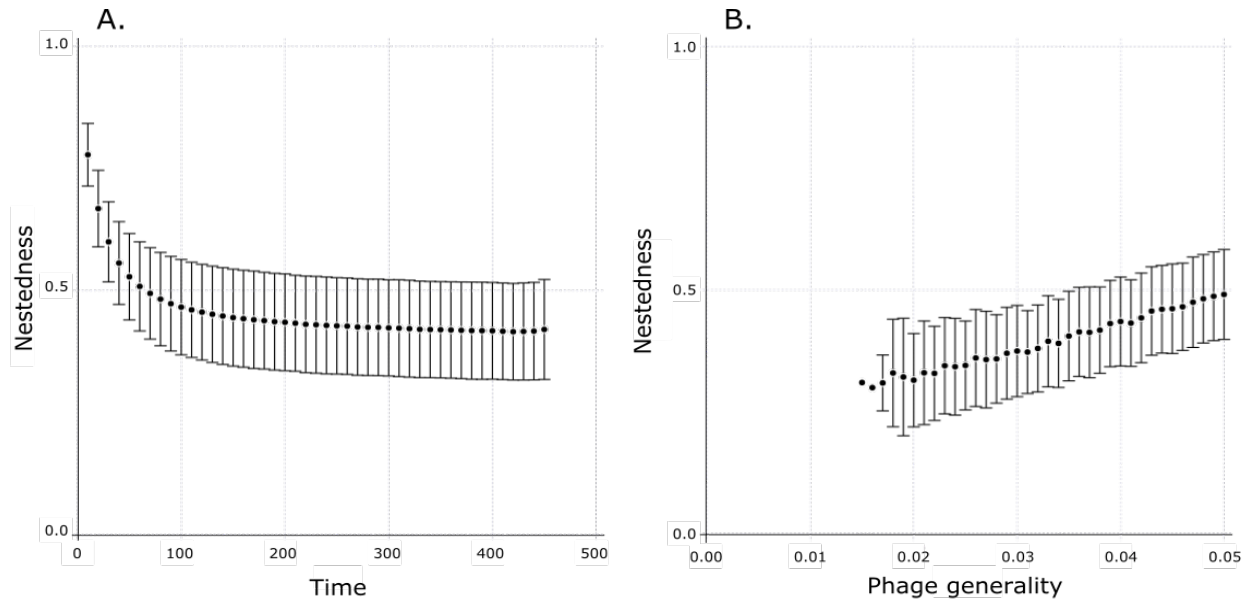
Figure 5.4 Evolution of mean nestedness. We selected simulations that last at least 450 time steps to investigate mean nestedness over time (A.) and mean nestedness depending on phage generality (B.). Bars represent standard deviation.

5.3, we thus have a structure that indicates a slowing down of dynamics with time. However, according to Fortuna et al. (2019), the nested structure would be similar to a fluctuating selection dynamic. The generality of the phages still plays an important role in the structuring of the networks. Indeed, we notice a difference in nestedness between specialist phages (0.30, $\pm$ 0.057) and generalists (0.49, $\pm$ 0.092). We could not note any influence of the virulence of the phages on the structure of the interactions (see additional figures).

5.4   Discussion

In this paper, we wanted to compare network structure to a time-shift experiment to evaluate its efficiency as an evolutionary dynamic indicator. With our model, we were able to rebuild evolving phage-bacteria populations. The time-shift experiments showed an ARD, while nestedness indicates FSD. As we also wanted to see the influence of phages traits on coevolution, we highlighted that phage virulence did not affect the evolution of the network structure. On the contrary, the generality of the phages influences the structure, with more nested networks

117

when the phages are more generalized.

Starting with the relationship between phage traits and the structure of interaction networks, the fact that the presence of generalist phages induces a more nested network structure can be explained by several things. Simulations including more generalist phages would show a higher probability of infection due to the higher interaction range, leading to a type of evolution similar to the gene-for-gene model, which is associated with a nested structure. The fact that very specific interactions between a parasite and its host produce instability and stochasticity in the system (Best et al. 2017) may also explain that simulations including generalist phages tended to last longer, as shown in figure 5.1 A. That is consistent with experimental results showing that generalist phages tend to emerge with long-term coevolution (Hall et al. 2011b). In our model, the generality of phages was fixed and did not evolve during simulations. The evolution between phage and bacteria and the structure of their interactions were not related to the evolution of generality but are related by phage traits fixed at the beginning of each simulation. We noticed that in all cases, the nestedness decreases with time, which can be associated with an adaptation of phages to bacteria over time, until a stabilization. This is also an evolution that has been found by Fortuna et al. (2019), where nestedness of antagonistic interactions tends to be higher with time under ARD and lower under FSD. The decrease of coevolutionary dynamic magnitude in our results could also emerge from the coevolution process and the trait matching between phages and bacteria (de Andreazzi et al. 2020).This can have been reinforced by the fact that we keep only simulations that ran long enough. Similarly, the trade-off between virulence and generality that appears in figure 5.1 A reflects the trade-off that has been highlighted in host-parasite systems, namely that virulence tends to be relatively low in parasites so that the infected population can continue to grow, allowing the parasite to persist (Buckling & Brockhurst 2008).

The time-shift experiments and network structure point to different dynamics within the same systems. Since the time-shift experiment is a proven method for coevolutionary dynamics

(Hall et al. 2020), it is difficult to imagine that the results obtained with these experiments could not be considered as valid. On top of that, bacteria and phages in experiments and nature usually show an ARD (Hall et al. 2011c). The network structure, on the other hand, has been little used as an indicator of coevolutionary dynamics, so its reliability cannot be guaranteed in this case. However, the two chosen indicators agree on one thing, our simulations do not present the change of dynamics observed in living environments. This indicates that the conflict we observe here lies more with our model than with the indicators themselves. To solve this particular issue, we could improve to way interactions between phages and bacteria are modeled.

In conclusion, the model presented here does not exactly reflect the reality as it appears in nature or in the laboratory, with specifically an absence of change in dynamics, but it remains an interesting tool, especially with regard to the factors influencing the structure of interactions. We have also been able to highlight that the structure of the interaction networks could not, in our case, be used in a certain and independent way if one wishes to obtain information concerning the coevolutionary dynamics between bacteria and phages. However, this model may be interesting to use to address the question of the impact of trait diversity on the speed, dynamics, and presence of coevolution, since the diversity of organisms interacting with bacteria hinders their evolution (Gómez & Buckling 2013), which could link with laboratory experiments and empirical studies.

Chapitre 6

Influence of interactions on community evolution, impact on network and phylogenetic structure

---

Influence of interactions on community evolution, impact on network and phylogenetic structure

Mathilde Besson[1,2], Dominique Gravel[2,6], Laura A. Burkle[3], Giulio V. Dalla Riva[4], Marcus de Aguiar[12], Marie-Joséee Fortin[5], Paulo R. Guimarães Jr.[7], David H. Hembry[8], Flávia M. Marquitti[12], Erica A. Newman[9,10], Jens M. Olesen[11], Mathias M. Pires[12], Justin D. Yeakel[13,14] & Timothée Poisot[1,2]

[1] Département de Sciences Biologiques, Université de Montréal, Montréal, H2V 2J7, Canada; [2] Québec Centre for Biodiversity Sciences, McGill University, Montréal, H3A 1B1, Canada; [3] Department of Ecology, Montana State University, Bozeman, MT 59715, U.S.A.; [4] Beaty Biodiversity Research Centre, University of British Columbia, Vancouver, V6T 1Z4, Canada; [5] Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, M5S 3B2, Canada; [6] Département de Biologie, Université de Sherbrooke, Sherbrooke, J1K 2R1, Canada; [7] Departamento de Ecologia, Instituto de Biociências, Universidade de São Paulo, São Paulo, 05508-090, Brazil; [8] Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, U.S.A.; [9] School of Natural Resources and Environment, University of Arizona, Tucson, AZ 85721, U.S.A.; [10] Pacific Wildland Fire Sciences Laboratory, USDA Forest Service, Seattle, WA 98103, U.S.A.; [11] Department of Bioscience, Aarhus University, Aarhus, 8000, Denmark; [12] Departamento de Biologia Animal, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Campinas, 13083-862, Brazil; [13] Life and Environmental Sciences, University of California Merced, Merced, CA 95343, U.S.A.; [14] Santa Fe Institute, Santa Fe, NM 87501, U.S.A

in prep.

Abstract

Species interactions are known to play a major role in the community evolution, in terms of dynamic and composition, and vice-versa. In evolutionary studies of communities, species richness is widely used to find episodes of diversifications. They can also be detected by the analysis of the phylogenetic tree, using for example the alpha statistic. The alpha statistic measures phylogenetic tree branch length and compares them to each other. The results of this comparison indicate whether diversification takes place early or later in the history of the community. Consideri5sng interactions, the structure of interactions network differs depending on the interaction type, *e.g.* competition or mutualism. This structure can be used as another indicator of community evolution. In this paper, we want to see if interaction type has a direct influence on community evolution and if it is possible to find an insight of these interactions in community history. We especially focus on positive and negative interactions, *i.e.* competition and facilitation. To do so, we build a community model based on species traits allowing us to track phylogenetic tree structure over time and rebuild interactions network at any time, in which we can decide what will be the unique interaction type, positive or negative. To highlight differences in the community evolution, we calculated the alpha statistic, which help to detect moments of diversification in the community history. We then measured the network structure over time using modularity. We used an analysis of tree similarity for the phylogenetic trees. We found that interaction type does impact communities dynamic by changing the speed of community assembly, but it seems that this impact is no longer measurable after the community has reached a stable dynamic. Interaction type also impacts the phylogenetic tree structure as they appeared to be organized according to the speed of community building.

6.1   Introduction

Species interactions –defined here as any possible interaction between two or more species such as predation, competition or mutualism– are considered as one of the most important

factors influencing community evolution, in the dynamic and composition of communities other time and space. Beyond intraspecific competition, interspecific interactions have long been included in theoretical studies to understand the evolution of species and communities. For instance, Ehrlich & Raven (1964) included interactions inside their escape and radiate model. They highlighted the importance of trophic interactions on diversification rate (*i.e.* a noticeable change in the species richness of a community over time, related to changes in speciation and extinction rates), especially for plant/herbivore relationships. Few years later Van Valen (1973) developed his *Red Queen hypothesis* in which interactions, such as competition or predation, push species to adapt to the ever-changing environment and to the ever-evolving other species with which they interact. More recently, Thompson (2005) based his *Geographic Mosaic Theory of Coevolution* on spatial and/or temporal variation of interactions to explain different evolutionary outcomes between populations of the same species. This resulted in a landscape composed of hot evolutionary spots (*i.e.* ecological sites of high interaction and evolution intensities) surrounded by cold evolutionary spots (*i.e.* sites with lower interactions and slower evolution). These studies have improved our understanding of community evolution but where often restricted to only few species. How we apply these results to larger communities remains a challenge.

Interaction networks make a convenient formalism for the investigation of the evolution of community structure. Numerical models have proven to be useful tools because of the difficulty to study evolution by observations and experiments (Hall et al. 2020). They show how non-random structures of species organization arise from basic ecological principles. Models such as those of Williams & Martinez (2000) or Bastolla et al. (2005) open this field tracking how traits drive the structure of interaction networks. Such models were found efficient to mimic the properties of empirical networks, but they nonetheless lack basic principles of evolutionary dynamics. Loeuille & Loreau (2009) are among the first to present an evolutionary model for diverse communities of interacting species. Still based on traits, they can reconstruct

species interaction networks with an architecture similar to real networks (Brännström et al. 2012). The most important finding of these studies is that basic evolutionary principles combined with a simple representation of network structure allows to recreate and track evolutionary trajectories of entire communities. Knowledge acquired from these models can not however be directly translated to other types of interactions networks and therefore we still to investigate the influence of competition and mutualism on the macro evolution of communities.

Several approaches can be used to follow community evolution. Metrics reporting species richness or phylogenetic trees inform on changes in origination, extinction or diversification rates over time and point out key moments in the evolutionary history. For instance, Quental & Marshall (2011) used measures of branch length to infer the diversity dynamics of a community. A more significant challenge is to relate the evolutionary history to the development and change of network structure. As closely related species tend to have similar interacting species (Rafferty & Ives 2013 in mutualistic networks), Cattin et al. (2004b) showed that phylogenetic constraints were partly responsible for the structure of interaction networks, while Peralta et al. (2015) showed that the link between phylogeny and networks was dependent on the trophic levels observed. It is possible to recreate interaction networks based on phylogenics trees (Peralta 2016a). But revealing quantitatively how the two types of structures relate is not always straightforward. These differences may be induced by the close link between the structure of interaction networks and the species richness of the observed communities (Chomicki et al. 2019). As discussed above, studies related to the link between species richness and network structure and phylogenetic trees are generally restricted to food webs and plant communities. The relationship between them is still blurry in different types of communities.

In communities studies, the evolution of species richness is widely used as an indicator of evolution. Species dynamics in a community provide information on its evolution and serve as a basis for measuring it. Based on that, indices such as speciation, extinction or diversification rates, are used to quantify this evolution, by the evaluation of their changes over time. It helps

to detect key moments in the evolutionary history of communities. This information about species richness dynamics can also be extracted from phylogenetic trees. A variety of indices have been developed to quantify this evolution, for instance by defining a diversification rate of a community. Among these, the phylogenetic tree structure is a good proxy of community evolution. Measures such as phylogenetic tree balance or the distribution of branches length within a tree give us clues about how evolution took place. Branch length distribution informs when diversification events appeared in communities' history. Species richness indices are also used with interaction network studies, but usually, it's the other way around. Indeed, species interactions influence species richness, as Chomicki et al. (2019) showed in their review on the impact of mutualism on species richness.

Species interaction networks provide a comprehensive snapshot of community structure at a given time in its history. The structure of the network has a direct impact on the evolutionary dynamics, in particular it influences traits involved in the occurrence of these interactions (Guimarães et al. 2017). The type of interaction that is taken into account is also important in the outcome of evolution. Positive (*i.e.* mutualism or facilitation) and negative (*i.e.* competition, parasitism or predation) interactions will have different impact on community structure. For instance, negative interaction appears to add a strong selective pressure on species, with a large impact on their fitness causing traits displacement (Silvestro et al. 2015) and increasing diversification rates. The influence of positive interactions is however still unclear. First, the number of studies focussing on positive interaction is much lower than the one investigating negative interactions influence on evolution. Then, depending on the model used or the species you look at, it is possible to have contrasting results. We know that interaction types lead to particular evolutionary output, but the quantity of effort on the different interaction types is unbalanced and need more investigation for positive interaction network to be able to reach a consensus.

In reverse, evolutionary dynamics obviously influence the structure of interactions networks

(Segar et al. 2020). On the one hand, species interaction networks of positive interactions, such as mutualistic interactions, tend to have a nested structure (a network in which specialist species tend to interact with a subgroup of species a little more generalist Bascompte et al. 2003). On the other hand, the networks of negative or antagonistic interactions tend to be more modular (subgroups of species that interact more with each other than with the rest of the species present in the community Newman & Girvan 2004) (Thébault & Fontaine 2010b). These two salient features of such networks are presumably the result of evolutionary dynamics (de Andreazzi et al. 2018). This reciprocal influence allows us to use measurements of networks structure to investigate macroevolution, and would also be useful to acknowledge a possible footprint of interactions in community evolution. As we know that interaction networks evolve over time (Peralta 2016a), it makes them an interesting tool to analyze and better understand communities' evolution. But the information on how they do evolve over long periods of time is still lacking (Poisot & Stouffer 2016). Despite the importance of interactions in evolutionary theory, their imprint on communities' evolution is not always detectable and straightforward to decrypt. Most species interactions are weak and rapidly fluctuate over time. Based on these facts, two conclusions can be formulated. First, because interactions vary largely and their strength is most of the time very weak, their imprint vanishes in a few generations. But, even if interactions are weak, their impact at one time in the community can be so strong that it will mark it and will strongly influence the evolution of this community. Even if there is no consensus on the fact that we can observe or not the signature of interactions, their implication on macro-evolution direction can not be contested.

Here, we focus on unipatite community networks, with negative or positive interactions. Negative interaction communities are guided by competition. For positive interactions, as we use unipatite communities, we decided to investigate facilitation. Facilitation is defined as an interaction in which the presence of one species modifies the environment, which promotes the growth, survival or reproduction of another nearby species (Bronstein 2009). Most studies

investigating facilitation focus on plant-plant communities studies (for example see verd08nap, Schöb et al. 2018), but it is possible to find few studies with other types of interaction such as trophic interactions focal (Bracken et al. 2007). Few studies investigated the relationship between facilitation and community evolution (see Bronstein 2009 ; Brooker et al. 2008), which gives us plenty of exploration axes. The interactions between species are inherent to the species present at a t time, their evolution should also be largely influenced by the species composing the community. Then, communities' evolution should be different according to the type of interaction taken into account, with faster community building in positive interaction communities than in negative ones. Turnover should be higher in competitive/negative interaction communities which can lead to a higher diversification rate than in facilitation communities.

In this paper, we want to evaluate if positive and negative interactions leave a distinct imprint on diversification. We focus on their joint impacts on phylogenetic and network structures, two elements of community structure that are susceptible to inform us on the reciprocal feedback between evolutionary dynamics and ecological interactions. For this purpose, we built a community model based on species traits. We have reconstructed the phylogeny of the entire community and species interaction network at any time and analyzed the alpha value indices to characterize the change in phylogenetic tree structure over time. We also looked at phylogenetic trees' similarities. Finally, we measured species interaction network structure via modularity over time. As expected, the two types of interactions lead to different community structures in term of phylogeny and network. We also found an organization of phylogenetic structure according to the speed of community building.

6.2 Methods

6.2.1 Model

Our model combines a niche model and evolutionary dynamics. For the interaction network, we focus on unipartite communities with positive or negative interactions (Figure 6.1). A single interaction type is assigned per community. We track long-term community evolution with a time step corresponding to a generation. Diversification rate is defined by origination and extinction rates, both of which are influenced by interspecific interactions. The niche space community is defined by a single axis with values ranging between zero and one. Each species is defined by traits, which influence interactions with the community and both traits are subject to mutations and evolution. We track for each newly established species the position into the niche space in order to rebuild its interactions with other species, the time of speciation and of extinction. This information allows us to rebuild phylogenies for every simulated community, the possible interactions between each species and the evolution of their traits.

Model functioning Each species of the community is characterized by two values that we call traits: the niche position/optimum n and the interaction range r (Figure 6.1). Both traits are ranging between 0 and 1 and we impose no correlation between them. The niche optimum indicates the position of each species in the niche space as well as its optimum. The interaction range provides an area inside which each present species will interact with the focal species, in other terms, its generality. The distribution of niche positions and ranges of all species from a community determine the entire network structure.

Each community start with one unique species for which the position on the niche space is picked randomly between 0 and 1. Mutations are driving speciation events and the emergence of new species. The new species, defined as its *parent* with two numerical traits, will successfully establish if interspecific interactions are favorable. The arrival of new species (*i.e.* origination

probability) is therefore conditioned by two probabilities. First, test (i) the probability of a speciation event, which depends on the number of already present spceies, and (ii) the probability of new formed species to establish in the community, which depends on the possible interactions. Once a speciation event occurs, meaning that the community does not contain too many species, traits of the descendents (n_{new} and r_{new}) are inherited from the ancestor with some variation, which we call mutation. n_{new} is randomly picked from a normal distribution which have the parent niche position as mean and a fixed variation (see table 6.1, sd). The same is done for with a variance of 0.2.

Establishment is then calculated, depending on the number of interactions (*i.e.* the degree) the potential new species would have with the resident community. More specifically, it is defined for positive or negative interactions are defined by the following equations:

$$P(establishment_{pos}) = u_{0pos} + u_{1pos}.e^{-a_u.I_m},$$

$$P(establishment_{neg}) = u_{0neg} + u_{1neg}.e^{-a_{uneg}.I_m},$$

where I_m is the sum of interactions a species have at a specific time step, u_0 and u_1 defined how fast and high the establishment probability will increase or dicrease, depending on the interaction type.

If a speciation event occures, the parent species will be remove for the community, leaving space for two new possible species (Figure 6.1 b. step 3 and Possibilities). This way of modeling evolution mimics cladogenese evolutionary procesus.

The algorithm follows with the extinction process, with extinction probability also a function of the number of interactions. More specifically, it is defined for positive or negative interactions

are defined by the following equations:

$$P(extinction_pos) = e_{0pos} + e_{1pos}.e^{-a_{epos}.I_m},$$

$$P(extinction_neg) = e_{0neg}.(1 - e^{-a_{eneg}.I_m}).$$

where I_m is the sum of interactions a species has at a specific time step, e_0 and e_1 defined how fast and high the establishment probability will increase or decrease, depending on the interaction type.

A simulation starts with a single ancestral species and stops when it has run for a certain amount of time steps, or when a maximal number of species (S_{max}) in the community has been reached.

6.2.2 Simulations

For each interaction type, positive or negative, 100 simulations were performed. This amount was sufficient to catch relevant information from the output as a higher number of simulations didn't reveal a significant change in the harvested information. Only simulations with communities composed of more than 20 species at the end of the simulation were kept for analysis. Four outputs were extracted for each run: (i) species present at each time step, (ii) traits values of each species (n and r), (iii) establishment time of each parent and child species, and (iv) extinction time of each species.

This model has 21 parameters. Some are common for the two types of interactions, some are specific to one or the other (Table 6.1). Shared parameters were, for instance, the maximal amount of interactions per species (I_{max}). It allows us to avoid logistic explosion of species richness in the communities, which saturates the model and makes the simulations endless. The total number of created species during each simulation (S_{max}) is also a way of avoiding

131
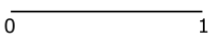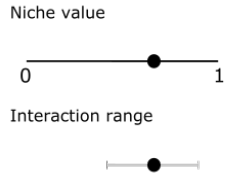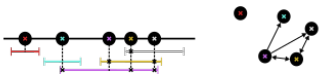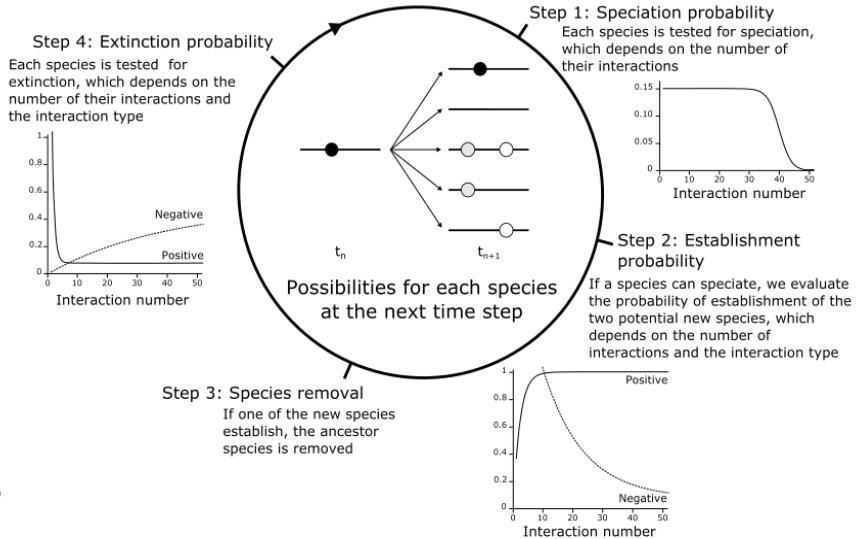
Figure 6.1 Conceptual illustration of the model used for the simulations. (a) The niche space is defined as a linear and continuous axis between 0 and 1 (a1). Each species is defined by two numerical values: the niche position/optimum which is a number between 0 and 1 and the interaction range, centered on the niche position (a2). An interaction occurs if the niche position of a species is located in the interaction range of another one (a3). Once that is set, the model can run following the same frame over and over (b). First, each species is tested for origination (speciation and establishment of new species)(steps 1 to 3). Second, they are tested for extinction (step 4). From the time n to the time n+1, there are 5 different possible situations: 1. the species remains, 2. it goes extinct, 3. it speciates and the two news species establish, 4. and 5. it speciates but only one new species establishes.

132

endless simulations. It has been set at 1000 species per simulation, which is largely enough to have the emergence of stable communities in term of species richness.

The speciation probability (u_{max}) is also common to all simulations/interaction types. It is the same for every species and fixed in time at 0.15, which is around 1 speciation each 7 time step. Concerning the extinction and establishment probabilities, parameters allow us to shape curves of these probability in a way that we can obtain stable communities. Details of these parameters are in table 6.1. Parameters related to these probabilities were determinant for the success of simulations.

Table 6.1 Model's parameters, signification and values.

Parameter	Signification	Value
int	Interaction type, 0 for competition, 1 for facilitation	
S_{max}	Maximal number of species in the system	1000
av_r	Half range of the niche of the first species	0.1
sd	Standard deviation of the normal distribution used to calculate the niche optimum trait of a child species	$2*av_r + 0.0001$
u_{max}	Speciation probability	0.1
a_{eneg}	Shape of the exponential decay of the negative extinction - interaction relationship	0.025
e_{0neg}	Asymptotic extinction probability with infinite negative interactions	0.5
e_{1neg}	Extinction probability with absence of competition interaction	1
a_{epos}	Shape of the exponential decay of the positive extinction - interaction relationship	1.2
e_{0pos}	Asymptotic extinction probability with infinite positive interactions	0.075
e_{1pos}	Extinction probability with absence of facilitation interaction	5.19
a_{uneg}	Shape of the exponential decay of the colonization - interaction relationship	0.075
u_{0neg}	Asymptotic establishment probability with infinite competition interactions	0.075
u_{1neg}	Establishment probability with absence of competition interaction	2
a_u	Shape of the exponential decay of the colonization - interaction relationship	0.45
u_{0pos}	Asymptotic establishment probability with infinite facilitation interactions	1

u_{1pos}	Establishment probability with absence of facilitation interactions	-1
d	Decrease speed of the establishment probability	0.5
B_{spe}	Constant minimal number of interaction per species (establishment prob)	4
B_{ext}	Constant minimal number of interaction per species (extinction prob)	4
I_{max}	Maximal number of interacting species	40

6.2.3 Analysis

We first recorded diversification rates over time. We followed the species richness dynamic of every simulation and calculated the mean species richness dynamic and its 95% confidence interval. Species turnover of each community over time was calculated using the `turnover` function of `codyn` R package.

We investigated variation in phylogenetic structure over time and between interaction types. We built phylogenies for each simulation using speciation time. We then calculated the α-value at each time step, for every community. The α-value is a measure of diversification rate, based on the γ statistic, which allows us to compare phylogenetic trees. It compares the branch length at the top and the bottom of a tree. If top branches are relatively longer than bottom ones, α is higher than 0. This means that the diversification process occurred recently in the community history. The opposite, α lower than 0, means that diversification was faster at the beginning of the community history. For the analysis, we built bifurcation trees with the information on parentality and extinction time.

We recorded the degree, k, of each species at each time step (i.e. the number of interaction a species have). We also calculated network modularity once communities reached a relativelly stable number of species. To do so, each community networks were divided into sub-groups of highly connected species using random walk technique. We then tested how good this sub-groups division was using the `modularity` function of `igraph` R package.

We wanted to verify if there was a particular structuration of communities regarding phylogenetic and network structure.

Phylogenies at the end of simulations (when species number reaches S_{max}) were used for comparison. In this way, all trees have the same number of tips and can be compared. Topography similarity was then compared and visualized using `treespace` function from `Treespace` R package. This analysis performs an ordination of trees similarity using a principal coordinate analysis (PCoA or MSD), based on a distance matrix between trees. This distance

matrix is made with trees topography branches length and structure according Kendall and Colijn metric. We applied a hierarchical clustering on the previous PCoA using the Ward's method.

For both analysis, in order to highlight a particular structure in topography and network, communities have been aggregated into groups. To assign each community to a group, we looked at their velocity to reach species richness stability. As a proxy of this velocity, we used the time to reach the mipoint of the mean species richness over time curve. Groups are formed by communities with similar velocity. They are made by regular intervals between the fastest and the slowest community. In regards of velocity distribution, we decided to create 5 groups. Group size depends on the velocity similarity and will not be the same for each of them.

6.3   Results

When we look at community diversification dynamics, it shows a significantly variation within the same interaction during the radiation phase. But most communities have reached an equilibrium of species richness after roughly 150 time steps. We therefore considered this time step as a reference to calculate indices for stable communities. As we can see in figure 6.2 A, the speed to reach this stable state, represented by the slope of the mean species richness evolution, differs from positive to negative interactions communities. In average, positive interaction communities diversify earlier and faster than negative interaction communities. At steady state, we note that negative interactions generate a wider range of variation in species richness. Both interaction types exhibit a convergence of origination and extinction rates with increasing species richness (figure 6.2 B). Finally, species turnover at steady state is also more important for negative interactions (6.2 C).

We then compared the phylogenetic structures among the two interaction types once communities are stable. The α statistic tends to be relatively the same for positive and negative interactions once communites have reached a stable state, around 0 (figure 6.3 A), meaning
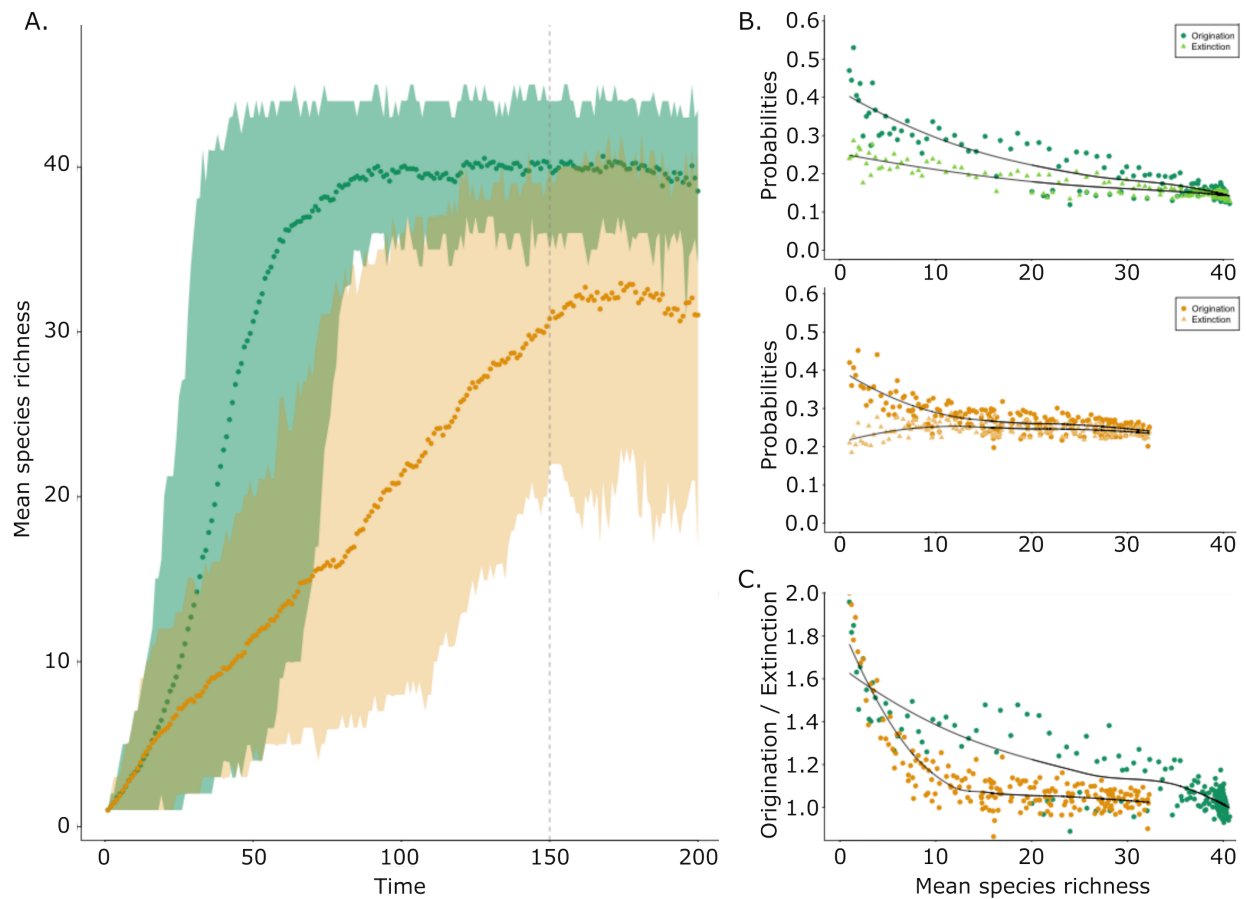
Figure 6.2 Diversification dynamics. Positive interactions are represented in green and negative interactions in orange. Light color areas around means denote the 95% of respective distributions. (A) Evolution of species richness over time. (B) Relationship between species richness and origination (plain circle) and extinction (plain triangle) rates, green color for positive interaction, orange for negative ones. (C) Relationship between diversification rate and species richness. The origination:extinction ratio indicates the species turnover within communities.

that, at this time, the branch length over the trees are quite similar, indicating no particular diversification point. At this point, variation within communities is slightly larger for negative than for positive interactions. However, when considering the history of communities, the shape of the α statistic indicates a shift in communities dynamics. This shift happens in average at timestep 17 for negative interactions communities and timestep 42 for positive interaction communities (figure 6.3 A). Before this point, new species struggle to establish within communities due to the few numbers of already established species and constraints imposed by origination and extinction probabilities. After this point establishment within communities become easier, which induce an increase of the α statistic. This is reinforced by the result of figure 6.3 B, showing the evolution og the α statistic according to species richness and time. We see with this figure that α statistic, so diversification rate, starts to increase when the community is composed by a certain minimum of species (around *5.26* species for negative interaction communities and around *24.68* species for positive interactions communities).

Interaction type also impacts the evolution of network structure. As we can see in figure 6.4 A, at stable state, mean species degree (*i.e.* mean number of interactions per species) is larger for positive interactions, with species interacting on average with *32.34* species, than for species from negative interactions communities interact with *7.30* species in average. Most importantly, the mean degree follows species richness for positive interaction communities while it quickly saturates largely before species richness stabilizes within negative interactions communities (figure 6.2 A). Variation of the mean degree also differs, being larger for positive interactions communities than negative ones. The evolution of the mean degree is mostly driven by selection on a smaller range (*0.056* for negative interactions communities against *0.254* for positive ones). Mean modularity also differs depending on interaction type (figure 6.4 B). Negative interactions communities network are more modular than positive ones, with mean modularity up to *0.593* for the first one and *0.032* for the second. The evolution of
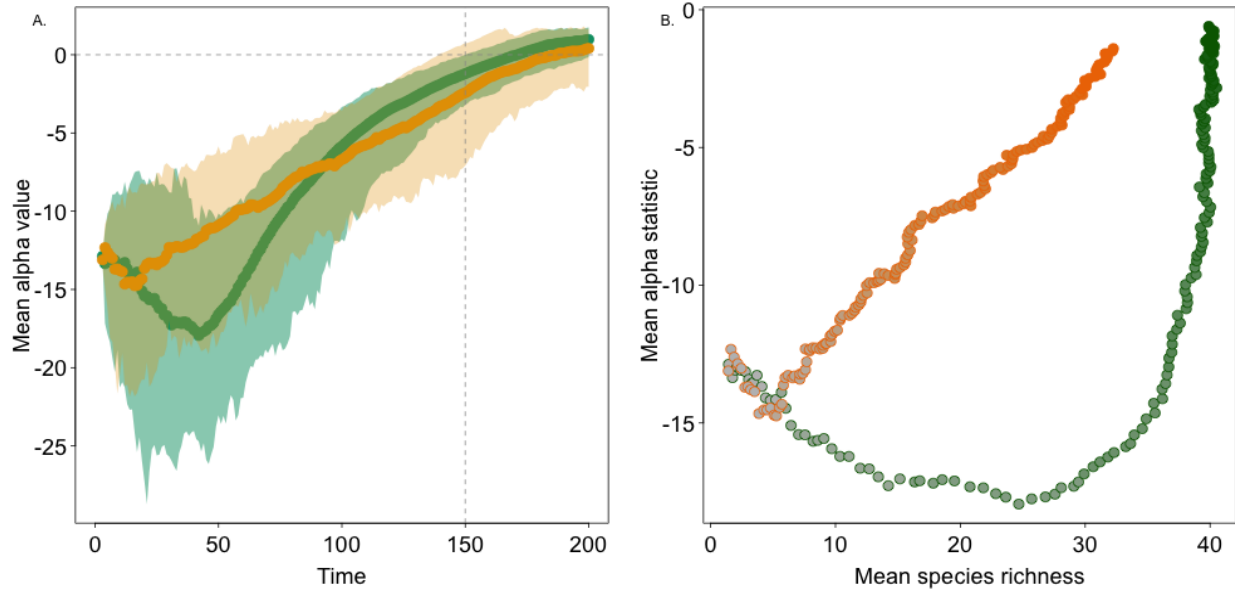
Figure 6.3 Evolution of phylogenetic structure. Green represents indices calculated for positive interaction communities, orange is used to illustrate indices calculated for negative interaction communities. Light colors areas around means include 95% of communities. (A) Evolution of α over time. (B) α statistic according to species richness and time. Grey to color scale represents time flow, grey is the beginning of communities. The dot color increases with time.

mean modularity is dependent on community species richness, with two different trajectories depending on the interaction type. Modularity follows the evolution of species richness in negative interactions communities while it tends to stabilize around zero in positive interactions communities. On the one hand, selective pressure forces species of negative interactions communities to interact with a few amount of species, creating over time and communities building, small groups of interacting species. On the other, species from positive interactions communities are not subject to the same pressure, which is reflected by random organization of species interactions.

Lastly, we investigated if phylogeny and network structure jointly reflect community building history. A PCoA on phylogenetic tree similarity shows that the evolutionary history differ from positive and negative interactions. Trees from positive and negative interaction communities are separated into two slightly overlapping groups (figure 6.5 A). There is a larger variation within phylogenetic tree structure in negative interactions communities than within positive ones. For
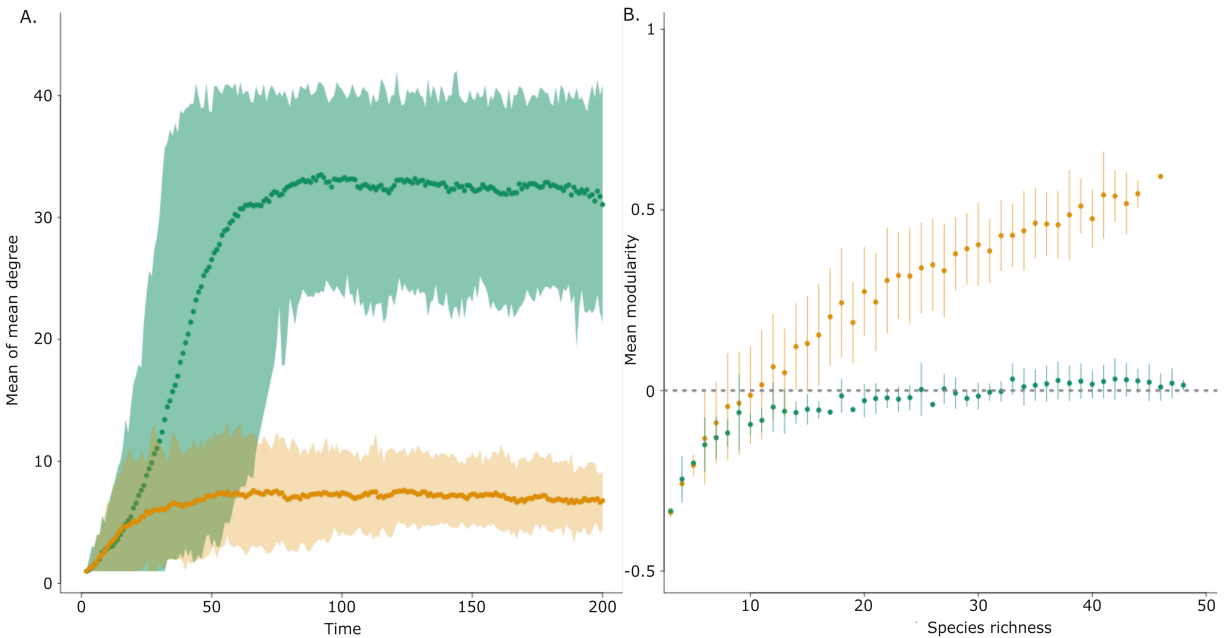
Figure 6.4 Network evolution and interaction types. Green represents indices calculated for positive interaction. (A) Evolution of mean species degree over time. Light colors areas around means include 95% of communities. (B) Relationship between species richness and communities modularity. Vertical bars represent the associated standard deviation of each mean.

each interaction type, communities have been divided into groups depending on the inner group similarity. The hierarchical clustering gave four groups of similar tree structure within each interactions type (figure 6.5 B). In both cases, each group is related to the community velocity to reach a stable number of species, from faster to slower communities. This last figure shows that phylogenetic tree structures are affected by the history of communities, independently of interactions type. The indices we used to describe our communities evolution (*i.e.* alpha value or network modularity) do not show a particular organisation according to the groups.

6.4 Discussion

With this paper, we wanted to see if interactions leaves a particular insight depending on their type. To investigate that, we followed assigned a unique interaction type to each commnities and tracked the evolution of species richness, network structure and phylogenetic trees structure.
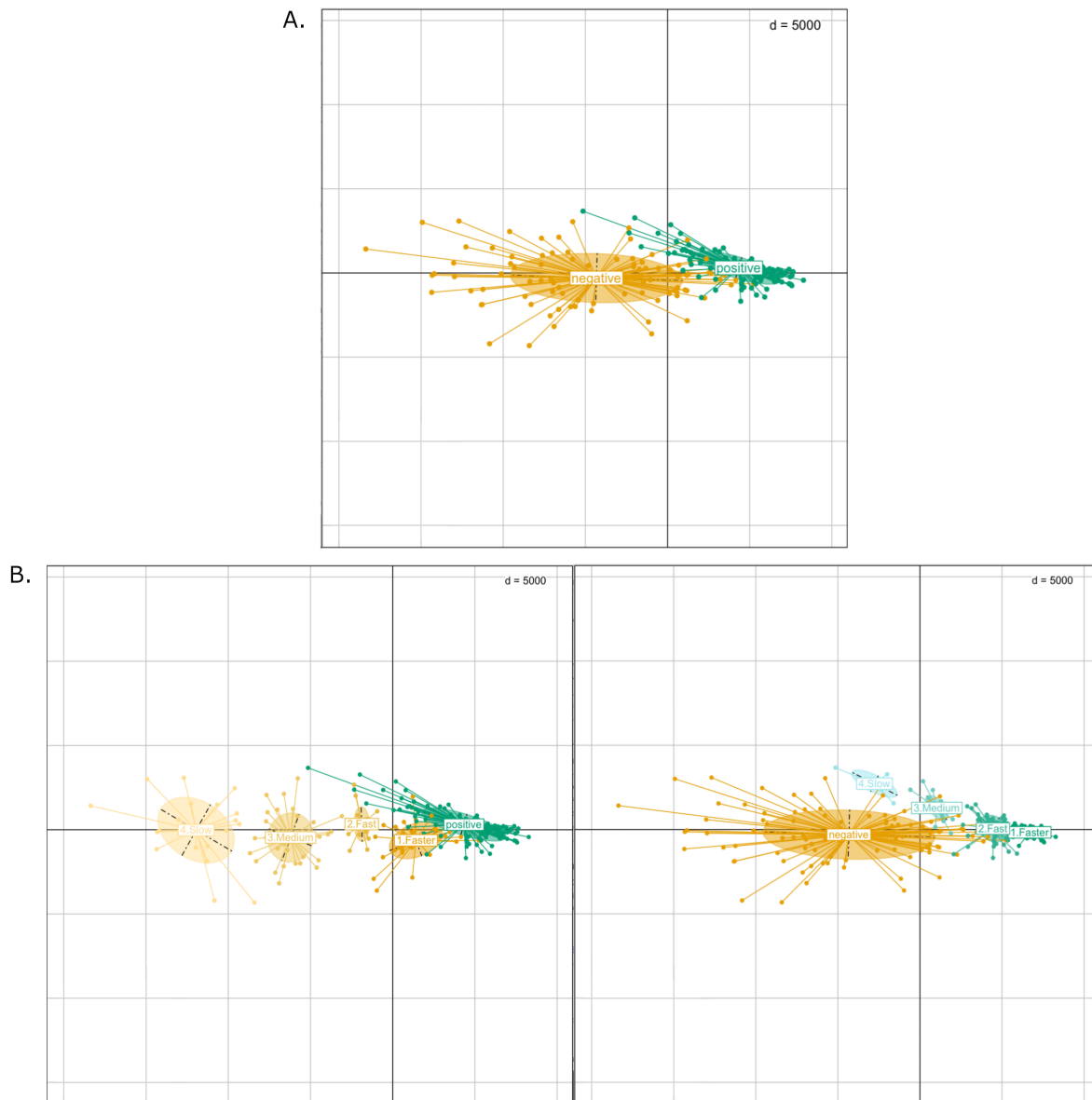
Figure 6.5 Ordination of phylogenetic tree structure with PCoA. Positive interactions communities are represented in green, negative interactions communities in orange. Each dot represents a phylogentic tree of a community. (A) Phylogenetic trees can be separate into two distinct groups, according to the interaction type of communities. (B) Within both of interaction type, groups are well organized according to their community building history.

We have found some insight, but not exactly in the way we tought. Our results shown that the structure of phylogenetic trees of communities is organized according to the type of interactions, but also according to the evolutionary speed of these communities. Species richness plays a determining role in the evolution of the structure of phylogenetic trees, regardless of the type of interactions considered. The relationship between interaction network modularity and species richness is also important for negative interaction communities. However, the link between species richness and the evolution of interaction network structure in general remains unclear.

The differences observed in communities dynamics (6.2) are consistent with other studies founding. For instance, Schluter & Pennell (2017) highlighted that species was a major determinant of diversification. They found higher rate of speciation when diversity was low, caused by the opulence of species free niches. These results support the Diversity Begets Diversity theory, where diversification is higher in low community diversity (Madi et al. 2020). All of this confirms the fact that our model is able to recreate reality like community dymanics and gives us confiance on the further results. With regard to interaction type in diversification rate (6.2 C), our results show that, as Zeng & Wiens (2021), positive interactions communities tend to have a higher diversification rate than negative ones.This can also be linked to the fact that positive interactions species tend to have more interactions (degree) than species from negative interaction communities, as we can observe for microbiome communities (Estrela et al. 2022).

The evolution of the phylogenetic trees structure obtained with our model (figure 6.3 A) corresponds to what has already been shown previously with the α value (Marquitti et al. 2020) or its equivalent the γ value (Quental & Marshall 2011). Our results show that this evolution is dependent on the species richness of the community. Figure 6.3 B shows that this dependency is stronger for the communities of negative interactions than for the communities of positive interactions, where a shift in the average alpha value occurs when the average species richness of a community reaches 10 species. For the positive interactions communities, this shift is more

gradual, which shows that the selection pressure certainly due to the constraints of interactions are stronger for the communities of negative interactions. In all cases, all communities tend towards an alpha value close to zero, regardless of their interaction type and their speed of reaching a stable number of species. However, the effectiveness of α statistic like indicators to test diversification based on phylogenetical tree are not as trustable as we thought (Louca & Pennell 2020). In this case, maybe the PCoA on phylogenetic trees can gives more valuable insight. The results of our PCoA show that the trees is organized according to the speed of reaching the equilibrium in the number of species (figure 6.5), but also according to the interaction type, which is also fibd by Segar et al. (2020). The structuration may be caused by the time communities have been at the equilibrium when we measured the phylogenetic trees structure. The final observed structure difference could then be explained by species degree.

The evolution of the network interaction structure also corresponds to what one would expect depending on the type of interaction considered. Negative interaction networks become more and more modular over time while positive interaction networks do not show this structure, which is consistent with what have has been shown on the subjet. The seminal paper by Thébault & Fontaine (2010b) showed indeed the differences in structure depending on the type of interaction (*i.e.* mutualistic networks are nested while antagonistic network are modular). The fact that we find a neutral structure for the positive interaction networks is also explained by the number of species in our communities, and consistent with Maliet et al. (2020) results. Olesen et al. (2007) also showed that mutualistic communities with less than 50 species did not show any particular modular structure. In figure 6.4, we see that species richness is driving modularity of negative interactions communities, because of the reduction of niche availability (Ezard & Purvis 2016). Unfortunately, we were not able to highlight a link between the evolution of network structure and community dynamics with simple descriptors of network structure. To be able to relate the evolution of the network structure and the specific richness of the communities, the indices used were not sufficient. The comparison of networks and their

evolution is still a challenge today. The use of graphlets would be a possible way to go further in the investigation of network evolution (Pržulj 2007; Rahman et al. 2014 ; Hulovatyy et al. 2015).

In the light of these results, we can note that interactions play an major role in the formation of communities and that we can visualize some small differences in the establishment of communities, as well as community dynamics. However, in the long term it can be difficult to detect a significant difference in the structure of the communities when only looking at species richness. However, network structure and phylogenetic trees topologies tend to keed a lit insight of interaction type. As Burin et al. (2021) did in their paper, it would also be interesting to investigate the relationship between phylogeny and network structure. The type of selection pressure differs according to the type of interaction, but in the end the studied structures are quite similar. Moreover, in the empirical communities, the types of interactions are not clearly separated as they are in our model. Therefore, the distinction between the signatures of the different types of interactions should be all the more difficult to detect. This ability to distinctly separate interaction types is not something found in empirical communities or in the laboratory. However, it is useful for understanding the functioning and influence of each in the evolution of communities. There are studies that count many types of interactions, like in Melián et al. (2009), but it is a framework that is still needed in evolutionary ecology research (Fontaine et al. 2011a). So one of the next steps in this model could be to allow interactions to change over time, across the community. As Bronstein et al. (2003) have shown, the combination of positive and negative interactions (one antagonistic and two mutualistic species) can greatly influence the evolutionary dynamics of the traits of interacting species, the real challenge here would be to implement these combinations for more than a three-species system.

Chapitre 7

Conclusion

L'objectif principal de cette thèse était de renforcer les connaissances sur l'impact des interactions sur l'évolution des écosystèmes. Pour atteindre cet objectif, j'ai décidé de me focaliser sur plusieurs échelles d'observation, à savoir l'échelle de l'individu et des communautés, ce qui me parait indispensable pour avoir une meilleure compréhension de mécanismes en jeu et des liens entre ces niveaux d'organisation au niveau évolutif. Pour cela, j'ai utlisé les réseaux d'interactions.

L'approche des réseaux d'interactions est introduite dans la chapitre 2. Ce chapitre d'encyclopédie fait office d'introduction à l'utilisation des réseaux d'interactions en écologie. Il met de l'avant les principales avancées qu'a permis de faire l'approche par réseaux et les défis qui lui font face, mais également la possibilité que cette approche apporte dans la mise en place de nouvelles méthodes. Les méthodes développées avec l'approche par réseaux se retrouvent dans le chapitre 3 qui présente une revue, non exhaustive, des outils d'analyses utilisant les réseaux d'interactions en écologie, ainsi que leurs limitations. Les deux chapitres suivants se focalisent sur l'utilisation des réseaux dans l'étude des interactions antagonistes, et plus spécifiquement entre les bactéries et les phages. Le chapitre 4, qui à l'origine devait représenter la partie expérimentale de l'utilisation des réseaux dans un contexte évolutif, a malheureusement dû être abandonné. Il a cependant servi de base à l'élaboration du modèle utilisé pour le chapitre 5. Ce chapitre se concentre sur la relation entre les interactions antagonistes bactéries-phages et leur coévolution. Le modèle utilisé permet de reproduire les expérimentations en laboratoire, et pourrait possiblement aider à perfectionner celles-ci. La structure des réseaux, telle qu'elle est utilisée dans ce chapitre, ne permet pas de donner des indications claires sur la dynamique coévolutive entre les bactéries et leurs phages. Cependant, cela nous a permis de mettre en évidence que la généralité des phages jouait un rôle non négligeable dans la structure des réseaux d'interactions entre phages et bactéries. Enfin, le chapitre 6 considèrent les réseaux d'interactions à plus large échelle, celle de la communauté. Nous avons pu ainsi constater que le type d'interactions (positive ou négative) influence la façon dont les communautés d'espèce

s'établissent et évoluent au début de leur histoire évolutive. Cependant, le lien entre la structure des réseaux d'interactions et celle des arbres phylogénétiques n'a pas pu être clairement établi.

7.1 La structure des réseaux entre micro et macro

Comme nous avons pu le voir au long de cette thèse, les réseaux sont partout, à tous les niveaux et dans tous les systèmes. Grâce à l'étude de leur structure, ils peuvent aider à dresser un portrait de l'histoire évolutive (Proulx et al. 2005b), où chaque image retranscrit une part de l'information, complétant les connaissances déjà acquises via d'autres méthodes. Les réseaux offrent la possibilité de faire des suivis temporels aussi bien que spatiaux, même si le développement de la méthodologie est encore nécessaire aujourd'hui.

Dans les deux derniers chapitres, nous avons abordé l'utilisation des réseaux comme indicateur de l'évolution et ce à différentes échelles. Le chapitre 5 se focalisait sur l'évolution conjointe à l'échelle locale, alors que le chapitre 6 portait sur l'évolution des communautés, donc à une échelle un peu plus large. Dans les deux cas, la structure des réseaux d'interactions, entre populations ou entre espèces a été calculée. Comme nous avons pu le souligner dans les chapitre 2 et chapitre 3, il existe une multitude de mesures pour étudier la structure des réseaux. Dans les deux derniers chapitres, nous nous sommes cependant concentrés presque uniquement sur l'emboitement d'un côté et la modularité de l'autre. Dans les deux cas, nous avons constaté une évolution des réseaux au cours du temps, comme avait également pu le mettre en évidence Peralta (2016b). Cependant, la structure des réseaux n'était pas toujours un bon indicateur de l'évolution, peu importe les échelles que nous avons utilisées. Toujours est-il que la modularité comme l'emboitement semble fortement lié à la connectance et au nombre d'espèces et d'interactions [fort10nmea]. Bien que ce lien ait été vérifié pour les réseaux à large échelle, le lien avec la connectance gagnerait a être approfondi.

Malgré des différences apparentes dans les types de modèles utilisés (individus centrés pour l'étude de la coévolution et community level pour l'influence des interactions sur la dynamique

149

des communautés), les deux chapitres se rejoignent sur la façon de conceptualiser les traits. En effet, dans les deux cas, les individus et les espèces étaient caractérisées par deux valeurs numériques représentant deux traits : la position dans la niche et le range d'interactions. C'est peut-être ce point particulier qui permettra à d'autres d'apporter des réponses aux questions ouvertes ici et qui n'oa pas pu être répondue aussi clairement que désiré.

7.2 Utilisation des réseaux en évolution

L'un des points forts de l'approche par réseaux pour l'étude de l'évolution est que les mesures effectuées sur ceux-ci peuvent être appliquées à l'ensemble des niveaux d'organisation. Les différentes structures, emboitement et modularité, sont en effets observés à différents niveaux (Cantor et al. 2017). Cela montre que la structure est indépendante du niveau d'organisation, mais elle l'est également des échelles spatiales et temporelles (Guimarães 2020), ce qui rend l'étude des réseaux d'interactions particulièrement intéressante pour comprendre les systèmes écologiques. Grâce à cela, il est possible de suivre l'évolution des écosystèmes tout faisant le lien avec les changements propres à chaque échelle d'organisation (démographique, énergétique, etc.). En effet, les interactions et leurs changements vont venir affecter le fitness des individus, ce qui aura un impact à plus large échelle sur la démographie puis sur l'évolution des traits d'une espèce par exemple. Des changements au niveau individuel peuvent aussi avoir lieu sans pour autant que la structure des réseaux aux niveaux supérieurs ne change, elle aussi. Cependant, les structures observées au niveau des espèces viennent de la façon dont les individus interagissent entre eux. La présence de généralistes et de spécialistes a aussi une influence sur la structure des réseaux à l'échelle de l'espèce, rendant les réseaux plus ou moins connectés ou déconnecter. Cela a des conséquences sur l'évolution des communautés et des espèces puisque quand une espèce généraliste connecte plusieurs réseaux déconnectés, considérés alors comme un hub. Cela offre plus de chemins possibles dans les réseaux, et donc plus d'interactions possibles. Cependant, l'influence de ces propriétés sur des échelles plus grandes sont encore floues. Les

liens évolutifs entre les différents niveaux d'organisation sont certains, mais la connaissance à ce sujet est encore un peu fébrile. Et comme l'indique Guimarães (2020), le prochain grand défi pour l'approche par réseaux en évolution sera de réellement comprendre et interpréter correctement les similarités de structure de réseaux à différentes échelles d'observations.

À mesure que l'échelle d'observation augmente, le degré de complexité des réseaux s'accroît, dû à des réseaux plus grands car plus divers (Eklöf et al. 2013b). Cela peut rendre leur étude plus fastidieuse et leur compréhension plus difficile, et de surcroit rendre plus compliquer encore les liens entre les échelles. C'est ici que la modélisation théorique des réseaux a son importance. Les modèles adaptatifs par exemple apportent une partie du lien manquant entre les processus sous-jacents à l'évolution des traits et la structure des réseaux d'interactions au niveau des communautés (Gross & Sayama 2009). L'importance des modèles mathématiques dans l'avancée des connaissances en écologie et évolution n'est donc pas à sous-estimer. De plus, les modèles permettent d'accéder à de l'information qui serait autrement difficile d'accès et d'étude. C'est le cas des interactions d'ordre supérieur ou encore des interactions indirectes. En effet, Guimarães et al. (2017) ont pu montrer, grâce à un modèle basé sur les réseaux que les espèces qui n'interagissaient pas entre-elles directement pouvaient tout de même avoir une influence significative sur la coévolution entre mutualistes. Les modèles théoriques sont donc d'une aide précieuse tant dans la compréhension de l'évolution à travers l'étude des réseaux d'interactions, mais également et surtout pour construire des liens entre les différents niveaux d'organisation (Segar et al. 2020).

En conclusion, l'apport combiné des modèles mathématiques et des réseaux d'interactions a permis de faire des avancées majeures dans la compréhension des dynamiques évolutives qui façonnent les écosystèmes. Les réseaux d'interactions apparaissent comme étant un bon moyen de faire le lien entre les différentes échelles d'études en évolution. L'augmentation de collectes de données sur les interactions écologiques à différents niveaux, associées à l'augmentation de la puissance des modèles mathématiques appliqués à l'écologie et l'évolution, ne pourront

qu'aider à développer de nouvelles méthodes basées sur les réseaux qui nous permettront d'approfondir ce domaine, mais pour l'instant le manque de données disponibles est un facteur limitant important. Cette thèse s'insère donc parfaitement dans ce contexte en proposant des modèles mathématiques générant ces propres données et ce, à différentes échelles, mais tout en gardant la structure des réseaux d'interaction comme fil conducteur. Même si l'utilisation des réseaux pour étudier l'évolution n'est pas toujours concluante, l'angle d'approche utilisée pour tenter d'aider à la construction d'un cadre conceptuel pour l'étude de l'évolution par les réseaux, ouvre la porte à un univers de création sans limites. En effet, les modèles utilisés peuvent être facilement modifiés. C'est d'ailleurs de cas du modèle de coévolution qui donne la possibilité d'augmenter la diversité des espèces ou populations étudiées. Loin d'être une contribution de taille dans le domaine des réseaux et de l'évolution, cette thèse peut donc se prendre comme une invitation à aller chercher plus loin.

Références

Agrawal* & Lively. (2002). Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evol Ecol Res.* 4:91–107.

Albert & Barabási. (2002). Statistical mechanics of complex networks. *Rev Mod Phys.* 74:47–97.

Alexander et al. (2013). Trait-mediated indirect interactions in a marine intertidal system as quantified by functional responses. *Oikos.* 122:1521–31.

Allesina & Bodini. (2004). Who dominates whom in the ecosystem? Energy flow bottlenecks and cascading extinctions. *Journal of Theoretical Biology.* 230:351–8.

Allesina & Pascual. (2007). Network structure, predator–prey modules, and stability in large food webs. *Theor Ecol.* 1:55–64.

Allesina & Pascual. (2009a). Food web models: a plea for groups. *Ecology Letters.* 12:652–62.

Allesina & Pascual. (2009b). Googling Food Webs: Can an Eigenvector Measure Species' Importance for Coextinctions? *PLoS Comput Biol.* 5:e1000494.

Allesina & Tang. (2015). The stability–complexity relationship at age 40: a random matrix perspective. *Population Ecology.* 57:63–75.

Almeida-Neto et al. (2007). On nestedness analyses: rethinking matrix temperature and anti-nestedness. *Oikos.* 116:716–22.

Andreazzi et al. (2017). Network Structure and Selection Asymmetry Drive Coevolution in Species-Rich Antagonistic Interactions. *The American Naturalist.* The University of Chicago Press; 190:99–115.

Araújo et al. (2008). Network analysis reveals contrasting effects of intraspecific competition

on individual vs. population diets. *Ecology.* 89:1981–93.

Araújo & Luoto. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography.* 16:743–53.

Araújo & Rozenfeld. (2014). The geographic scaling of biotic interactions. *Ecography.* 37:406–15.

Araújo et al. (2011). Using species co-occurrence networks to assess the impacts of climate change. *Ecography.* 34:897–908.

Araújo et al. (2006). Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography.* 33:1712–28.

Argov et al. (2017). Temperate bacteriophages as regulators of host behavior. *Current Opinion in Microbiology.* 38:81–7.

Atmar & Patterson. (1993). The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia.* 96:373–82.

Bahlai & Landis. (2016). Predicting plant attractiveness to pollinators with passive crowdsourcing. *Royal Society Open Science.* 3:150677.

Baiser et al. (2011). Species Richness and Trophic Diversity Increase Decomposition in a Co-Evolved Food Web. *PLOS ONE.* 6:e20672.

Baiser et al. (2012). Geographic variation in network structure of a nearctic aquatic food web. *Global Ecology and Biogeography.* 21:579–91.

Baker et al. (2014). Species' roles in food webs show fidelity across a highly variable oak forest. *Ecography.* 38:130–9.

Baker et al. (2015). Species' roles in food webs show fidelity across a highly variable oak forest. *Ecography.* 38:130–9.

Banašek-Richter et al. (2004). Sampling effects and the robustness of quantitative and qualitative food-web descriptors. *Journal of Theoretical Biology.* 226:23–32.

Bartomeus. (2013). Understanding Linkage Rules in Plant-Pollinator Networks by Using

Hierarchical Models That Incorporate Pollinator Detectability and Plant Traits. *PLOS ONE.* Public Library of Science; 8:e69200.

Bartomeus et al. (2016). A common framework for identifying linkage rules across different types of interactions. *Functional Ecology.* 30:1894–903.

Bascompte. (2007). Networks in ecology. *Basic and Applied Ecology.* 8:485–90.

Bascompte et al. (2003). The nested assembly of plant–animal mutualistic networks. *PNAS.* National Academy of Sciences; 100:9383–7.

Bascompte et al. (2006). Asymmetric Coevolutionary Networks Facilitate Biodiversity Maintenance. *Science.* American Association for the Advancement of Science; 312:431–3.

Bascompte & Melián. (2005). Simple Trophic Modules for Complex Food Webs. *Ecology.* 86:2868–73.

Bascompte et al. (2005). Interaction strength combinations and the overfishing of a marine food web. *PNAS.* National Academy of Sciences; 102:5443–7.

Baskerville et al. (2011). Spatial Guilds in the Serengeti Food Web Revealed by a Bayesian Group Model. *PLoS Comput Biol.* 7:e1002321.

Bastolla et al. (2009). The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature.* Nature Publishing Group; 458:1018–20.

Bastolla et al. (2005). Biodiversity in model ecosystems, I: coexistence conditions for competing species. *Journal of Theoretical Biology.* 235:521–30.

Beckett & Williams. (2013). Coevolutionary diversification creates nested-modular structure in phage–bacteria interaction networks. *Interface Focus.* 3:20130033.

Bell et al. (1999). Centrality measures for disease transmission networks. *Social Networks.* 21:1–21.

Berlow et al. (2009). Simple prediction of interaction strengths in complex food webs. *Proc Natl Acad Sci U S A.* 106:187–91.

Bersier et al. (2002). Quantitative Descriptors of Food-Web Matrices. *Ecology.* 83:2394–407.

Best et al. (2017). Host–parasite fluctuating selection in the absence of specificity. *Proceedings of the Royal Society B: Biological Sciences.* Royal Society; 284:20171615.

Betts et al. (2014). Contrasted coevolutionary dynamics between a bacterial pathogen and its bacteriophages. *PNAS.* 111:11109–14.

Billick & Case. (1994). Higher Order Interactions in Ecological Communities: What Are They and How Can They be Detected? *Ecology.* 75:1529–43.

Blois et al. (2014). A framework for evaluating the influence of climate, dispersal limitation, and biotic interactions using fossil pollen associations across the late Quaternary. *Ecography.* 37:1095–108.

Blondel et al. (2008). Fast unfolding of communities in large networks. *J Stat Mech Theory Exp.* 2008:P10008.

Blüthgen et al. (2008). What do interaction network metrics tell us about specialization and biological traits? *Ecology.* 89:3387–99.

Bohannan & Lenski. (2000). Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. *Ecology Letters.* 3:362–77.

Bonacich. (1987). Power and Centrality: A Family of Measures. *American Journal of Sociology.* 92:1170–82.

Borrelli. (2015). Selection against instability: stable subgraphs are most frequent in empirical food webs. *Oikos.*:n/a–a.

Borrett et al. (2014). The rise of Network Ecology: Maps of the topic diversity and scientific collaboration. *Ecological Modelling.* 293:111–27.

Boulangeat et al. (2012). Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters.* 15:584–93.

Bracken et al. (2007). Whole-Community Mutualism: Associated Invertebrates Facilitate a Dominant Habitat-Forming Seaweed. *Ecology.* 88:2211–9.

Brännström et al. (2012). Modelling the ecology and evolution of communities: a review

of past achievements, current efforts, and future promises. *Evol Ecol Res.* Evolutionary Ecology, Ltd. 14:601–25.

Breitbart & Rohwer. (2005). Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology.* 13:278–84.

Brockhurst et al. (2003a). Population mixing accelerates coevolution. *Ecology Letters.* 6:975–9.

Brockhurst et al. (2003b). Population mixing accelerates coevolution. *Ecology Letters.* 6:975–9.

Bronstein. (2009). The evolution of facilitation and mutualism. *Journal of Ecology.* 97:1160–70.

Bronstein et al. (2003). Ecological dynamics of mutualist/antagonist communities. *Am Nat.* 162:S24–39.

Brooker et al. (2008). Facilitation in plant communities: the past, the present, and the future. *Journal of Ecology.* 96:18–34.

Brose et al. (2006a). Consumer–Resource Body-Size Relationships in Natural Food Webs. *Ecology.* 87:2411–7.

Brose et al. (2004). Unified spatial scaling of species and their trophic interactions. *Nature.* Nature Publishing Group; 428:167–71.

Brose et al. (2006b). Allometric scaling enhances stability in complex food webs. *Ecology Letters.* 9:1228–36.

Buckling & Brockhurst. (2008). Kin selection and the evolution of virulence. *Heredity.* Nature Publishing Group; 100:484–8.

Buckling & Brockhurst. (2012). Bacteria–Virus Coevolution. In: Soyer, ed. *Evolutionary Systems Biology.* New York, NY: Springer; pp. 347–70.

Buckling & Rainey. (2002). Antagonistic coevolution between a bacterium and a bacterio-phage. *Proceedings of the Royal Society of London Series B: Biological Sciences.* Royal Society; 269:931–6.

Buckling et al. (2006). Antagonistic coevolution with parasites increases the cost of host deleterious mutations. *Proceedings of the Royal Society B: Biological Sciences.* 273:45–9.

Bull. (1994). Virulence. *Evolution.* 48:1423–37.

Burin et al. (2021). Macroevolutionary stability predicts interaction patterns of species in seed dispersal networks. *Science.* American Association for the Advancement of Science; 372:733–7.

Butler & Ellison. (2007). Nitrogen cycling dynamics in the carnivorous northern pitcher plant, Sarracenia purpurea. *Functional Ecology.* 21:835–43.

Cadotte & Davies. (2016). Phylogenies in ecology: a guide to concepts and methods. Princeton: Princeton University Press;

Caldarelli. (2007). Scale-free networks: complex webs in nature and technology. *OUP Cat.*

Camacho et al. (2007). Quantitative analysis of the local structure of food webs. *Journal of Theoretical Biology.* 246:260–8.

Camerano. (1880). Dell'equilibrio dei viventi merce la reciproca distruzione. *Atti Della R Accad Delle Sci Torino.* 15:393–414.

Campbell et al. (2011). A network model for plant–pollinator community assembly. *PNAS.* National Academy of Sciences; 108:197–202.

Canard et al. (2012). Emergence of Structural Patterns in Neutral Trophic Networks. *PLOS ONE.* Public Library of Science; 7:e38295.

Canard et al. (2014). Empirical Evaluation of Neutral Interactions in Host-Parasite Networks. *The American Naturalist.* The University of Chicago Press; 183:468–79.

Cantor et al. (2017). Nestedness across biological scales. *PLOS ONE.* Public Library of Science; 12:e0171691.

Capitán et al. (2009). Statistical Mechanics of Ecosystem Assembly. *Phys Rev Lett.* American Physical Society; 103:168101.

Carmona et al. (2015). Fifty years of co-evolution and beyond: integrating co-evolution from

molecules to species. *Molecular Ecology.* 24:5315–29.

Carscallen et al. (2012). Estimating trophic position in marine and estuarine food webs. *Ecosphere.* 3:art25.

Carstensen et al. (2014). Beta Diversity of Plant-Pollinator Networks and the Spatial Turnover of Pairwise Interactions. *PLOS ONE.* Public Library of Science; 9:e112903.

Cattin et al. (2004a). Phylogenetic constraints and adaptation explain food-web structure. *Nature.* Nature Publishing Group; 427:835–9.

Cattin et al. (2004b). Phylogenetic constraints and adaptation explain food-web structure. *Nature.* Nature Publishing Group; 427:835–9.

Cazelles et al. (2015). A theory for species co-occurrence in interaction networks. *Theoretical Ecology.* 9:39–48.

Cazelles et al. (2016). A theory for species co-occurrence in interaction networks. *Theor Ecol.* 9:39–48.

Chacoff et al. (2012). Evaluating sampling completeness in a desert plant-pollinator network. *J Anim Ecol.* 81:190–200.

Chadès et al. (2011). General rules for managing and surveying networks of pests, diseases, and endangered species. *Proceedings of the National Academy of Sciences.* 108:8323–8.

Chagnon. (2015). Characterizing topology of ecological networks along gradients: The limits of metrics' standardization. *Ecological Complexity.* 22:36–9.

Chao et al. (2005). A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters.* 8:148–59.

Chesson. (2000). Mechanisms of Maintenance of Species Diversity. *Annu Rev Ecol Syst.* Annual Reviews; 31:343–66.

Chesson & Kuang. (2008). The interaction between predation and competition. *Nature.* 456:235–8.

Chomicki et al. (2019). The Impact of Mutualisms on Species Richness. *Trends in Ecology &*

Evolution. 34:698–711.

Cochran-Stafira & von Ende. (1998). Integrating Bacteria into Food Webs: Studies with Sarracenia Purpurea Inquilines. *Ecology.* 79:880–98.

Cohen & Briand. (1984). Trophic links of community food webs. *Proc Natl Acad Sci U S A.* 81:4105–9.

Coux et al. (2016). Linking species functional roles to their network roles. Mouillot, ed. *Ecology Letters.*

Crea et al. (2015). A new model for ecological networks using species-level traits. *Methods Ecol Evol.* 7:232–41.

Crea et al. (2016). A new model for ecological networks using species-level traits. *Methods in Ecology and Evolution.* 7:232–41.

Csardi & Nepusz. (2006). The igraph Software Package for Complex Network Research. *InterJournal.* Complex Systems:1695.

Dale & Fortin. (2010). From Graphs to Spatial Graphs. *Annual Review of Ecology, Evolution, and Systematics.* 41:21–38.

Dalla Riva & Stouffer. (2015). Exploring the evolutionary signature of food webs' backbones using functional traits. *Oikos.* 125:446–56.

Dalsgaard et al. (2013). Historical climate-change influences modularity and nestedness of pollination networks. *Ecography.* 36:1331–40.

Darwin et al. (1859). On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life [Internet]. London : John Murray, Albemarle Street, pp. 1–564.

Davis et al. (1998). Individualistic species responses invalidate simple physiological models of community dynamics under global environmental change. *Journal of Animal Ecology.* 67:600–12.

D. C. Lopez-Pascua & Buckling. (2008). Increasing productivity accelerates host–parasite

coevolution. *Journal of Evolutionary Biology.* 21:853–60.

de Andreazzi et al. (2020). Coevolution by different functional mechanisms modulates the structure and dynamics of antagonistic and mutualistic networks. *Oikos.* 129:224–37.

de Andreazzi et al. (2018). Eco-evolutionary feedbacks promote fluctuating selection and long-term stability of antagonistic networks. *Proceedings of the Royal Society B: Biological Sciences.* Royal Society; 285:20172596.

Delmas et al. (2019). Analysing ecological networks of species interactions. *Biological Reviews.* 94:16–36.

Desjardins-Proulx et al. (2017). Ecological interactions and the Netflix problem. *PeerJ.* PeerJ Inc. 5:e3644.

Devictor et al. (2010). Defining and measuring ecological specialization. *Journal of Applied Ecology.* 47:15–25.

Díaz-Castelazo et al. (2010). Changes of a mutualistic network over time: reanalysis over a 10-year period. *Ecology.* Ecological Society of America; 91:793–801.

Donohue et al. (2013). On the dimensionality of ecological stability. *Ecology Letters.* 16:421–9.

Dormann et al. (2017). Identifying Causes of Patterns in Ecological Networks: Opportunities and Limitations. *Annual Review of Ecology, Evolution, and Systematics.* 48:559–84.

Dormann et al. (2008). Introducing the bipartite Package: Analysing Ecological Networks. *R News.* 8:8–11.

Drake. (1991). Community-Assembly Mechanics and the Structure of an Experimental Species Ensemble. *The American Naturalist.* The University of Chicago Press; 137:1–26.

Duffy et al. (2007). The functional role of biodiversity in ecosystems: incorporating trophic complexity. *Ecology Letters.* 10:522–38.

Dunne. (2006). The network structure of food webs. *Ecological Networks: Linking Structure to Dynamics in Food Webs.* Oxford University Press, USA;

Dunne et al. (2002a). Food-web structure and network theory: The role of connectance and

size. *PNAS.* 99:12917–22.

Dunne et al. (2002b). Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology Letters.* 5:558–67.

Dupont et al. (2009). Spatio-temporal variation in the structure of pollination networks. *Oikos.* 118:1261–9.

Dupont et al. (2014). Spatial structure of an individual-based plant-pollinator network. *Oikos.* 123:1301–10.

Dupont et al. (2011). Scaling down from species to individuals: a flower–visitation network between individual honeybees and thistle plants. *Oikos.* 120:170–7.

Earn. (2000). Coherence and Conservation. *Science.* 290:1360–4.

Egerton. (2002). A History of the Ecological Sciences, Part 6: Arabic Language Science: Origins and Zoological Writings. *Bulletin of the Ecological Society of America.* 83.

Ehrlich & Raven. (1964). BUTTERFLIES AND PLANTS: A STUDY IN COEVOLUTION. *Evolution.* 18:586–608.

Eklöf et al. (2013a). The dimensionality of ecological networks. *Ecol Lett.* 16:577–83.

Eklöf et al. (2013b). The dimensionality of ecological networks. *Ecology Letters.* 16:577–83.

Eklöf & Stouffer. (2015). The phylogenetic component of food web structure and intervality. *Theoretical Ecology.*

Elton. (1927). Animal Ecology. London: Sidgwick and Jackson;

Elton. (1958). The Reasons for Conservation. In: Elton, ed. *The Ecology of Invasions by Animals and Plants.* Dordrecht: Springer Netherlands; pp. 143–53.

Emmerson & Raffaelli. (2004). Predator–prey body size, interaction strength and the stability of a real food web. *Journal of Animal Ecology.* 73:399–409.

Estrada & Bodin. (2008). Using network centrality measures to manage landscape connectivity. *Ecological Applications.* 18:1810–25.

Estrela et al. 2022 Feb. Diversity begets diversity under microbial niche construction [Internet].

Ecology;

Evans et al. (2016). Merging DNA metabarcoding and ecological network analysis to understand and build resilient terrestrial ecosystems. *Functional Ecology.*

Ezard & Purvis. (2016). Environmental changes define ecological limits to species richness and reveal the mode of macroevolutionary competition. *Ecology Letters.* 19:899–906.

Farage et al. (2020). A dynamical perspective to community detection in ecological networks. *bioRxiv.* Cold Spring Harbor Laboratory;2020.04.14.040519.

Faust & Skvoretz. (2002). Comparing Networks across Space and Time, Size and Species. *Sociological Methodology.* 32:267–99.

Flor. (1956). The Complementary Genic Systems in Flax and Flax Rust**Joint contribution from the Field Crops Research Branch, Agricultural Research Service, United States Department of Agriculture and the North Dakota Agricultural Experiment Station. *Advances in Genetics.* Elsevier; pp. 29–54.

Flores et al. (2011). Statistical structure of host–phage interactions. *PNAS.* 108:E288–97.

Flores et al. (2016). BiMAT: a MATLAB package to facilitate the analysis and visualization of bipartite networks. *Methods in Ecology and Evolution.* 7:127–32.

Flores et al. (2013). Multi-scale structure and geographic drivers of cross-infection within marine bacteria and phages. *ISME J.* 7:520–32.

Fontaine et al. (2011a). The ecological and evolutionary implications of merging different types of networks. *Ecology Letters.* 14:1170–81.

Fontaine et al. (2011b). The ecological and evolutionary implications of merging different types of networks. *Ecology Letters.* 14:1170–81.

Fontaine & Thébault. (2015). Comparing the conservatism of ecological interactions in plant–pollinator and plant–herbivore networks. *Population Ecology.* 57:29–36.

Fortuna et al. (2019). Coevolutionary dynamics shape the structure of bacteria-phage infection networks. *Evolution.* 73:1001–11.

Fortuna & Bascompte. (2006). Habitat loss and the structure of plant–animal mutualistic networks. *Ecology Letters.* 9:281–6.

Fortuna et al. (2010). Nestedness versus modularity in ecological networks: two sides of the same coin? *Journal of Animal Ecology.* 79:811–7.

Fox. (1988). Diffuse Coevolution Within Complex Communities. *Ecology.* 69:906–7.

Freeman. (1977). A Set of Measures of Centrality Based on Betweenness. *Sociometry.* 40:35–41.

Freeman. (1978). Centrality in social networks conceptual clarification. *Social Networks.* 1:215–39.

Freeman et al. (1979). Centrality in social networks: ii. experimental results. *Social Networks.* 2:119–41.

Friman & Buckling. (2013). Effects of predation on real-time host–parasite coevolutionary dynamics. *Ecol Lett.* 16:39–46.

Gaba & Ebert. (2009). Time-shift experiments as a tool to study antagonistic coevolution. *Trends in Ecology & Evolution.* 24:226–32.

Gandon et al. (2008). Host–parasite coevolution and patterns of adaptation across time and space. *Journal of Evolutionary Biology.* 21:1861–6.

Gauzens et al. (2015). Trophic groups and modules: two levels of group detection in food webs. *J R Soc Interface.* 12.

Golubski & Abrams. (2011). Modifying modifiers: what happens when interspecific interactions interact? *Journal of Animal Ecology.* 80:1097–108.

Gomulkiewicz et al. (2007). Dos and don'ts of testing the geographic mosaic theory of coevolution. *Heredity.* 98:249–58.

Gomulkiewicz et al. (2000). Hot Spots, Cold Spots, and the Geographic Mosaic Theory of Coevolution. *The American Naturalist.* 156:156–74.

González-Varo & Traveset. (2016). The Labile Limits of Forbidden Interactions. *Trends in*

Ecology & Evolution. 31:700–10.

Gotelli. (2000). Null model analysis of species co-occurrence patterns. *Ecology.* 81:2606–21.

Gotelli & Entsminger. (2003). Swap Algorithms in Null Model Analysis. *Ecology.* 84:532–5.

Gotelli et al. (2010). Macroecological signals of species interactions in the Danish avifauna. *PNAS.* National Academy of Sciences; 107:5030–5.

Gouhier et al. (2010). Synchrony and Stability of Food Webs in Metacommunities. *The American Naturalist.* The University of Chicago Press; 175:E16–34.

Gómez & Buckling. (2011). Bacteria-Phage Antagonistic Coevolution in Soil. *Science.* American Association for the Advancement of Science; 332:106–9.

Gómez & Buckling. (2013). Coevolution with phages does not influence the evolution of bacterial mutation rates in soil. *ISME J.* Nature Publishing Group; 7:2242–4.

Gravel et al. (2016a). The meaning of functional trait composition of food webs for ecosystem functioning. *Philosophical Transactions of the Royal Society B: Biological Sciences.* Royal Society; 371:20150268.

Gravel et al. (2011). Trophic theory of island biogeography. *Ecology Letters.* 14:1010–6.

Gravel et al. (2016b). Stability and complexity in model meta-ecosystems. *Nature Communications.* 7:12457.

Gravel et al. (2013). Inferring food web structure from predator–prey body size relationships. *Methods Ecol Evol.* 4:1083–90.

Gross & Sayama. (2009). Adaptive Networks. In: Gross, Sayama, eds. *Adaptive Networks: Theory, Models and Applications.* Berlin, Heidelberg: Springer; pp. 1–8.

Guimarães. (2020). The Structure of Ecological Networks Across Levels of Organization. *Annual Review of Ecology, Evolution, and Systematics.* 51:433–60.

Guimarães et al. (2017). Indirect effects drive coevolution in mutualistic networks. *Nature.* Nature Publishing Group; 550:511–4.

Guimerà & Amaral. (2005). Cartography of complex networks: modules and universal roles.

Journal of Statistical Mechanics: Theory and Experiment. 2005:P02001.

Guimerà & Nunes Amaral. (2005). Functional cartography of complex metabolic networks. *Nature.* 433:895–900.

Guimerà et al. (2004). Modularity from fluctuations in random graphs and complex networks. *Physical Review E.* 70.

Gurney et al. (2017). Network structure and local adaptation in co-evolving bacteria–phage interactions. *Mol Ecol.*:n/a–a.

Hagberg et al. (2008). Exploring Network Structure, Dynamics, and Function using NetworkX. In: Varoquaux et al., eds. *Proceedings of the 7th Python in Science Conference.* Pasadena, CA USA; pp. 11–5.

Hall et al. (2020). Measuring Coevolutionary Dynamics in Species-Rich Communities. *Trends in Ecology & Evolution.* 35:539–50.

Hall et al. (2011a). Bacteria-Phage Coevolution and the Emergence of Generalist Pathogens. *The American Naturalist.* The University of Chicago Press; 177:44–53.

Hall et al. (2011b). Bacteria-Phage Coevolution and the Emergence of Generalist Pathogens. *The American Naturalist.* 177:44–53.

Hall et al. (2011c). Host–parasite coevolutionary arms races give way to fluctuating selection. *Ecology Letters.* 14:635–42.

Hall et al. (2011d). Host–parasite coevolutionary arms races give way to fluctuating selection. *Ecology Letters.* 14:635–42.

Harvey et al. (2013). Identity effects dominate the impacts of multiple species extinctions on the functioning of complex food webs. *Ecology.* 94:169–79.

Havens. (1992). Scale and Structure in Natural Food Webs. *Science.* American Association for the Advancement of Science; 257:1107–9.

Heikkinen et al. (2007). Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography.* 16:754–63.

Holland et al. (1983). Stochastic blockmodels: First steps. *Social Networks.* 5:109–37.

Holt. (1996). Food webs in space: an island biogeographic perspective. *Food webs.* Springer; pp. 313–23.

Holt. (1997). Community modules. In: Gange, Brown, eds. *Multitrophic Interactions in Terrestrial Systems.* Cambridge University Press; pp. 333–50.

Holt. (2009). Bringing the Hutchinsonian niche into the 21st century: Ecological and evolutionary perspectives. *Proceedings of the National Academy of Sciences.* 106:19659–65.

HUBBELL. (2001). The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32) [Internet]. Princeton University Press;

Hudson et al. (2013). Cheddar: analysis and visualisation of ecological communities in R. *Methods in Ecology and Evolution.* 4:99–104.

Hulovatyy et al. (2015). Exploring the structure and function of temporal networks with dynamic graphlets. *Bioinformatics.* 31:i171–80.

Hutchinson. (1961). The Paradox of the Plankton. *The American Naturalist.* The University of Chicago Press; 95:137–45.

Ingram & Vreeland. (1980). Differential effects of ethanol and hexanol on the Escherichia coli cell envelope. *J Bacteriol.* 144:481–8.

Jacquet et al. (2013). No complexity-stability relationship in natural communities. *arXiv preprint arXiv:13075364.*

Jacquet et al. (2016). No complexity–stability relationship in empirical ecosystems. *Nature Communications.* Nature Publishing Group; 7:12573.

Jacquet et al. (2017). Extensions of Island Biogeography Theory predict the scaling of functional trait composition with habitat area and isolation. *Ecology Letters.* 20:135–46.

Janzen. (1980). When is it coevolution. *Evolution.* 34:611–2.

Jordano. (2016). Sampling networks of ecological interactions. *Functional Ecology.* 30:1883–93.

Jordano & Bascompte. (2013). Mutualistic Networks [Internet]. Princeton Univ Press;

Jordano et al. (2003). Invariant properties in coevolutionary networks of plant–animal interactions. *Ecology Letters.* 6:69–81.

Jordán & Scheuring. (2004). Network ecology: topological constraints on ecosystem dynamics. *Phys Life Rev.* 1:139–72.

Katz. (1953). A new status index derived from sociometric analysis. *Psychometrika.* 18:39–43.

Kemp et al. (2017). Invariant antagonistic network structure despite high spatial and temporal turnover of interactions. *Ecography.* 40:1315–24.

Kéfi et al. (2012). More than a meal: integrating non-feeding interactions into food webs. *Ecol Lett.* 15:291–300.

Kim. (1993). The Competition Number and its Variants. In: John Gimbel, ed. *Annals of Discrete Mathematics.* Elsevier; pp. 313–26.

Kneitel & Miller. (2002). Resource and Top-Predator Regulation in the Pitcher Plant (sarracenia Purpurea) Inquiline Community. *Ecology.* 83:680–8.

Koleff et al. (2003). Measuring beta diversity for presence–absence data. *Journal of Animal Ecology.* 72:367–82.

Koskella. (2014). Bacteria-Phage Interactions across Time and Space: Merging Local Adaptation and Time-Shift Experiments to Understand Phage Evolution. *The American Naturalist.* The University of Chicago Press; 184:S9–S21.

Koskella & Brockhurst. (2014). Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev.* Oxford Academic; 38:916–31.

Krasnov et al. (2012). Phylogenetic Signal in Module Composition and Species Connectivity in Compartmentalized Host-Parasite Networks. *The American Naturalist.* The University of Chicago Press; 179:501–11.

Krasnov et al. (2004). Geographical variation in host specificity of fleas (Siphonaptera) parasitic on small mammals: the influence of phylogeny and local environmental conditions.

Ecography. 27:787–97.

Krause et al. (2003). Compartments revealed in food-web structure. *Nature.* 426:282–5.

Krishna et al. (2008). A neutral-niche theory of nestedness in mutualistic networks. *Oikos.* 117:1609–18.

Laliberté & Legendre. (2010). A distance-based framework for measuring functional diversity from multiple traits. *Ecology.* 91:299–305.

Legendre & Legendre. (2012). Numerical ecology. Third English edition. Oxford, UK: Elsevier;

Leggett et al. (2013). Experimental Evolution of Adaptive Phenotypic Plasticity in a Parasite. *Current Biology.* 23:139–42.

Lemos-Costa et al. (2015). Network analyses support the role of prey preferences in shaping resource use patterns within five animal populations. *Oikos.*:n/a–a.

Li et al. (2020). Network cross-validation by edge sampling. *arXiv:161204717 [stat].*

Liu & Murata. (2009). Community Detection in Large-Scale Bipartite Networks. *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology.* Institute of Electrical & Electronics Engineers (IEEE);

Loeuille & Loreau. (2005). Evolutionary emergence of size-structured food webs. *PNAS.* National Academy of Sciences; 102:5761–6.

Loeuille & Loreau. (2009). Emergence of complex food webstructure in community evolutionmodels. *Community Ecology.* Oxford University Press;

Loreau. (2010). Linking biodiversity and ecosystems: towards a unifying ecological theory. *Philosophical Transactions of the Royal Society B: Biological Sciences.* Royal Society; 365:49–60.

Louca & Pennell. (2020). Extant timetrees are consistent with a myriad of diversification histories. *Nature.* Nature Publishing Group; 580:502–5.

Lundberg et al. (2008). Species loss leads to community closure. *Ecology Letters.* 3:465–8.

MacArthur. (1955). Fluctuations of Animal Populations and a Measure of Community Stability.

Ecology. Ecological Society of America; 36:533–6.

MacArthur & Wilson. (1967). The Theory of Island Biogeography. Princeton University Press;

MacDonald et al. (2020). Revisiting the links-species scaling relationship in food webs. *bioRxiv.* Cold Spring Harbor Laboratory;2020.02.13.947531.

Madi et al. (2020). Does diversity beget diversity in microbiomes? Weigel et al., eds. *eLife.* eLife Sciences Publications, Ltd; 9:e58999.

Maliet et al. (2020). An individual-based model for the eco-evolutionary emergence of bipartite interaction networks. *Ecology Letters.* John Wiley & Sons, Ltd; 23:1623–34.

Malik et al. (2006). Comparative efficacy of ethanol and isopropanol against feline calicivirus, a norovirus surrogate. *American Journal of Infection Control.* 34:31–5.

Malthus. (1798). An Essay on the Principle of Population.:134.

Marquitti et al. (2020). Allopatry increases the balance of phylogenetic trees during radiation under neutral speciation. *Ecography.* 43:1487–98.

Martinez. (1992). Constant Connectance in Community Food Webs. *The American Naturalist.* [University of Chicago Press, American Society of Naturalists]; 139:1208–18.

Martinez et al. (1999). Effects of sampling effort on characterization of food-web structure. *Ecology.* 80:1044–55.

Martinez et al. (2006). Diversity, complexity, and persistence in large model ecosystems. In: Pascual, Dunne, eds. *Ecological Networks: Linking Structure to Dynamics in Food Webs.* Oxford ; New York: Oxford University Press, USA; pp. 163–85.

Martín González et al. (2010). Centrality measures and the importance of generalist species in pollination networks. *Ecological Complexity.* 7:36–43.

Massol et al. (2017). Chapter Four Island Biogeography of Food Webs. In: Bohan et al., eds. *Advances in Ecological Research.* Academic Press; pp. 183–262.

May. (1972). Will a Large Complex System be Stable? *Nature.* Nature Publishing Group;

238:413–4.

May. (1974). Stability and Complexity in Model Ecosystems. Princeton University Press;

Mayfield & Stouffer. (2017). Higher-order interactions capture unexplained complexity in diverse communities. *Nat Ecol Evol.* 1:0062.

McCann. (2000). The diversity–stability debate. *Nature.* Nature Publishing Group; 405:228–33.

McCann. (2007). Protecting biostructure. *Nature.* 446:29–9.

McCann. (2012). Food webs. Princeton, N.J: Princeton University Press;

McCann et al. (1998). Weak trophic interactions and the balance of nature. *Nature.* Nature Publishing Group; 395:794–8.

McDonald-Madden et al. (2016). Using food-web theory to conserve ecosystems. *Nat Commun.* 7:10245.

McGill et al. (2006). Rebuilding community ecology from functional traits. *Trends Ecol Evol.* 21:178–85.

Melián et al. (2014). Individual trait variation and diversity in food webs.

Melián et al. (2009). Diversity in a complex ecological network with two interaction types. *Oikos.* 118:122–30.

Memmott et al. (2007). Global warming and the disruption of plant-pollinator interactions. *Ecology Letters.* 10:710–7.

Memmott et al. (2004). Tolerance of pollination networks to species extinctions. *Proceedings of the Royal Society of London Series B: Biological Sciences.* Royal Society; 271:2605–11.

Middelboe et al. (2009). Bacteriophages drive strain diversification in a marine *Flavobacterium* : implications for phage resistance and physiological properties. *Environmental Microbiology.* 11:1971–82.

Milo et al. (2002). Network Motifs: Simple Building Blocks of Complex Networks. *Science.* American Association for the Advancement of Science; 298:824–7.

Minor et al. (2008). Uncertainty in spatially explicit population models. *Biological Conservation.*

141:956–70.

Mode. (1958). A MATHEMATICAL MODEL FOR THE CO-EVOLUTION OF OBLIGATE PARASITES AND THEIR HOSTS. *Evolution.* 12:158–65.

Monteiro & Faria. (2016). The interplay between population stability and food-web topology predicts the occurrence of motifs in complex food-webs. *Journal of Theoretical Biology.* 409:165–71.

Montoya et al. (2006). Ecological networks and their fragility. *Nature.* Nature Publishing Group; 442:259–64.

Montoya & Solé. (2002). Small World Patterns in Food Webs. *Journal of Theoretical Biology.* 214:405–12.

Montoya et al. (2009). Press perturbations and indirect effects in real food webs. *Ecology.* 90:2426–33.

Morales-Castilla et al. (2015). Inferring biotic interactions from proxies. *Trends in Ecology & Evolution.* 30:347–56.

Morand et al. (2002). Mammal density and patterns of ectoparasite species richness and abundance. *Oecologia.* 131:289–95.

Morin. (2011). Community ecology. Hoboken, N.J.: Wiley-Blackwell;

Morlon et al. (2014). Effects of trophic similarity on community composition. *Ecology Letters.* 17:1495–506.

Mouillot et al. (2013). A functional approach reveals community responses to disturbances. *Trends Ecol Evol.* 28:167–77.

Newman. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E.* 69.

Newman. (2010). Networks. An introduction [Internet]. New York, NY: Oxford University Press;

Newman & Girvan. (2004). Finding and evaluating community structure in networks. *Phys*

Rev E. American Physical Society; 69:026113.

Novak et al. (2016). Characterizing Species Interactions to Understand Press Perturbations: What Is the Community Matrix? *Annual Review of Ecology, Evolution, and Systematics.* 47:409–32.

Nuismer. (2006). Parasite Local Adaptation in a Geographic Mosaic. *Evolution.* 60:24–30.

O'Donnell et al. (2017). Spatial distribution of environmental DNA in a nearshore marine habitat. *PeerJ.* 5:e3044.

Olesen et al. (2011). Missing and forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences.* Royal Society; 278:725–32.

Olesen et al. (2007). The modularity of pollination networks. *PNAS.* National Academy of Sciences; 104:19891–6.

Olesen et al. (2008). Temporal Dynamics in a Pollination Network. *Ecology.* 89:1573–82.

Olesen et al. (2010). From Broadstone to Zackenberg: space, time and hierarchies in ecological networks. *Advances in Ecological Research.* 42:1.

Olito & Fox. (2015). Species traits and abundances predict metrics of plantpollinator network structure, but not pairwise interactions. *Oikos.* 124:428–36.

Paine. (1969). The Pisaster-Tegula Interaction: Prey Patches, Predator Food Preference, and Intertidal Community Structure. *Ecology.* 50:950–61.

Patten. (1978). Systems Approach to the Concept of Environment.

Pearson & Dawson. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography.* 12:361–71.

Pearson et al. (2002). SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. *Ecological Modelling.* 154:289–300.

Peralta. (2016a). Merging evolutionary history into species interaction networks. *Functional Ecology.* 30:1917–25.

Peralta. (2016b). Merging evolutionary history into species interaction networks. *Funct Ecol.*:n/a–a.

Peralta et al. (2015). Phylogenetic diversity and co-evolutionary signals among trophic levels change across a habitat edge. *Journal of Animal Ecology.* 84:364–72.

Petchey et al. (2008a). Size, foraging, and food web structure. *Proceedings of the National Academy of Sciences.* 105:4191–6.

Petchey et al. (2008b). Trophically Unique Species Are Vulnerable to Cascading Extinction. *The American Naturalist.* The University of Chicago Press; 171:568–79.

Petchey & Gaston. (2002). Functional diversity (FD), species richness and community composition. *Ecology Letters.* 5:402–11.

Piechnik et al. (2008). Food-web assembly during a classic biogeographic study: species' "trophic breadth" corresponds to colonization order. *Oikos.* 117:665–74.

Pilosof et al. (2015). The Multilayer Nature of Ecological Networks. *arXiv preprint arXiv:151104453.*

Pimm & Lawton. (1978). On feeding on more than one trophic level. *Nature.* 275:542–4.

Pires et al. (2015). Pleistocene megafaunal interaction networks became more vulnerable after human arrival. *Proceedings of the Royal Society B: Biological Sciences.* 282:20151367.

Poelen et al. (2014). Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecological Informatics.* 24:148–59.

Poisot. (2013). An a posteriori measure of network modularity. *F1000Research.* 2.

Poisot et al. (2015a). mangal – making ecological network analysis simple. *Ecography.*:n/a–a.

Poisot et al. (2012). The dissimilarity of species interaction networks. *Ecology Letters.* 15:1353–61.

Poisot et al. (2016a). The structure of probabilistic networks. Vamosi, ed. *Methods in Ecology and Evolution.* 7:303–12.

Poisot & Gravel. (2014). When is an ecological network complex? Connectance drives degree distribution and emerging network properties. *PeerJ.* PeerJ Inc. 2:e251.

Poisot et al. (2015b). Synthetic datasets and community tools for the rapid testing of ecological hypotheses. *Ecography.*

Poisot et al. (2017). Hosts, parasites and their interactions respond to different climatic variables. *Global Ecology and Biogeography.* 26:942–51.

Poisot et al. (2011). Resource availability affects the structure of a natural bacteria–bacteriophage community. *Biology Letters.* Royal Society; 7:201–4.

Poisot et al. (2013). Trophic complementarity drives the biodiversity–ecosystem functioning relationship in food webs. *Ecology Letters.* 16:853–61.

Poisot & Stouffer. (2016). How ecological networks evolve. *bioRxiv.* Cold Spring Harbor Laboratory;071993.

Poisot et al. (2015c). Beyond species: why ecological interaction networks vary through space and time. *Oikos.* 124:243–51.

Poisot et al. (2016b). Describe, understand and predict: why do we need networks in ecology? *Funct Ecol.* 30:1878–82.

Pollock et al. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution.* 5:397–406.

Ponisio et al. (2017). Opportunistic attachment assembles plant–pollinator networks. *Ecology Letters.* 20:1261–72.

Post. (2002). The long and short of food-chain length. *Trends Ecol Evol.* 17:269–77.

Poulin. (2010). Network analysis shining light on parasite ecology and diversity. *Trends in Parasitology.* 26:492–8.

Poullain et al. (2008). The Evolution of Specificity in Evolving and Coevolving Antagonistic Interactions Between a Bacteria and Its Phage. *Evolution.* 62:1–11.

Prankevicius & Cameron. (1991). Bacterial dinitrogen fixation in the leaf of the northern pitcher plant (Sarracenia purpurea). *Can J Bot.* 69:2296–8.

Proulx et al. (2005a). Network thinking in ecology and evolution. *Trends in Ecology &*

Evolution. 20:345–53.

Proulx et al. (2005b). Network thinking in ecology and evolution. *Trends in Ecology & Evolution*. 20:345–53.

Pržulj. (2007). Biological network comparison using graphlet degree distribution. *Bioinformatics*. 23:e177–83.

Quental & Marshall. (2011). The Molecular Phylogenetic Signature of Clades in Decline. *PLOS ONE*. Public Library of Science; 6:e25780.

Rafferty & Ives. (2013). Phylogenetic trait-based analyses of ecological networks. *Ecology*. 94:2321–33.

Rahman et al. (2014). Graft: An Efficient Graphlet Counting Method for Large Graph Analysis. *IEEE Transactions on Knowledge and Data Engineering*. 26:2466–78.

Ramos-Jiliberto et al. (2012). Topological plasticity increases robustness of mutualistic networks. *Journal of Animal Ecology*. 81:896–904.

Rezende et al. (2009). Compartments in a marine food web associated with phylogeny, body mass, and habitat structure. *Ecology Letters*. 12:779–88.

Rezende et al. (2007). Effects of phenotypic complementarity and phylogeny on the nested structure of mutualistic networks. *Oikos*. 116:1919–29.

Rodriguez-Girones & Santamaria. (2006). A new algorithm to calculate the nestedness temperature of presence-absence matrices. *Journal of Biogeography*. 33:924–35.

Roopnarine & Angielczyk. (2012). The evolutionary palaeoecology of species and the tragedy of the commons. *Biology Letters*. 8:147–50.

Rosvall et al. (2009). The map equation. *Eur Phys J Spec Top*. 178:13–23.

Rosvall & Bergstrom. (2008). Maps of random walks on complex networks reveal community structure. *PNAS*. National Academy of Sciences; 105:1118–23.

Sanfeliu & Fu. (1983). A distance measure between attributed relational graphs for pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*. SMC-13:353–62.

Santamaría & Rodríguez-Gironés. (2007). Linkage Rules for Plant–Pollinator Networks: Trait Complementarity or Exploitation Barriers? Chittka, ed. *PLoS Biology.* 5:e31.

Saramäki et al. (2007). Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E.* 75:027105.

Scanlan et al. (2015). Coevolution with Bacteriophages Drives Genome-Wide Host Evolution and Constrains the Acquisition of Abiotic-Beneficial Mutations. *Molecular Biology and Evolution.* 32:1425–35.

Scanlan et al. (2016). Parasite genetic distance and local adaptation in coevolving bacteria-bacteriophage populations. *Mol Ecol.*:n/a–a.

Schluter & Pennell. (2017). Speciation gradients and the distribution of biodiversity. *Nature.* 546:48–55.

Schöb et al. (2018). Evolution of facilitation requires diverse communities. *Nat Ecol Evol.* Nature Publishing Group; 2:1381–5.

Scofield et al. (2015). Potential changes in bacterial metabolism associated with increased water temperature and nutrient inputs in tropical humic lagoons. *Front Microbiol.* 6.

Segar et al. (2020). The Role of Evolution in Shaping Ecological Networks. *Trends in Ecology & Evolution.* 35:454–66.

Silvestro et al. (2015). The role of clade competition in the diversification of North American canids. *PNAS.* National Academy of Sciences; 112:8684–9.

Simberloff & Wilson. (1969). Experimental Zoogeography of Islands: The Colonization of Empty Islands. *Ecology.* 50:278–96.

Soffer & Vazquez. (2005). Network clustering coefficient without degree-correlation biases. *Physical Review E.* 71:057101.

Sokal & Rohlf. (1962). The comparison of dendrograms by objective methods. *Taxon.*:33–40.

Solé & Montoya. (2001). Complexity and fragility in ecological networks. *Proceedings of the Royal Society of London Series B: Biological Sciences.* Royal Society; 268:2039–45.

Staniczenko et al. (2013). The ghost of nestedness in ecological networks. *Nat Commun.* 4:1391.

Stock et al. (2017). Linear filtering reveals false negatives in species interaction data. *Scientific Reports.* Nature Publishing Group; 7:45908.

Stouffer & Bascompte. (2010). Understanding food-web persistence from local to global scales. *Ecology Letters.* 13:154–61.

Stouffer & Bascompte. (2011). Compartmentalization increases food-web persistence. *PNAS.* 108:3648–52.

Stouffer et al. (2006). A robust measure of food web intervality. *PNAS.* 103:19015–20.

Stouffer et al. (2007). Evidence for the existence of a robust pattern of prey selection in food webs. *Proceedings of the Royal Society B: Biological Sciences.* 274:1931–40.

Stouffer et al. (2012). Evolutionary Conservation of Species' Roles in Food Webs. *Science.* American Association for the Advancement of Science; 335:1489–92.

Strogatz. (2001). Exploring complex networks. *Nature.* 410:268–76.

Strong & Leroux. (2014). Impact of Non-Native Terrestrial Mammals on the Structure of the Terrestrial Mammal Food Web of Newfoundland, Canada. *PLOS ONE.* 9:e106264.

Stumpf et al. (2012). Statistical analysis of network data and evolution on GPUs: High-performance statistical computing. *Nature Precedings.*

Sugihara & Ye. (2009). Cooperative network dynamics. *Nature.* Nature Publishing Group; 458:979–80.

Suttle. (2005). Viruses in the sea. *Nature.* 437:356–61.

Thébault & Fontaine. (2010a). Stability of Ecological Communities and the Architecture of Mutualistic and Trophic Networks. *Science.* 329:853–6.

Thébault & Fontaine. (2010b). Stability of Ecological Communities and the Architecture of Mutualistic and Trophic Networks. *Science.* American Association for the Advancement of Science; 329:853–6.

Thébault et al. (2007). Cascading extinctions and ecosystem functioning: contrasting effects of diversity depending on food web structure. *Oikos.* 116:163–73.

Thébault & Loreau. (2003). Food-web constraints on biodiversity–ecosystem functioning relationships. *PNAS.* National Academy of Sciences; 100:14949–54.

Thompson. (2005). The Geographic Mosaic of Coevolution. University of Chicago Press;

Thompson et al. (2012). Food webs: reconciling the structure and function of biodiversity. *Trends in Ecology & Evolution.* 27:689–97.

Thompson & Townsend. (2000). Is resolution the solution?: the effect of taxonomic resolution on the calculated properties of three stream food webs. *Freshwater Biology.* 44:413–22.

Thrall & Burdon. (2003). Evolution of Virulence in a Plant Host-Pathogen Metapopulation. *Science.* 299:1735–7.

Trøjelsgaard et al. (2015). Geographical variation in mutualistic networks: similarity, turnover and partner fidelity. *Proceedings of the Royal Society B: Biological Sciences.* Royal Society; 282:20142925.

Trøjelsgaard & Olesen. (2016). Ecological networks in motion: micro- and macroscopic variability across scales. *Functional Ecology.* 30:1926–35.

Turcotte et al. (2012). Adaptive Evolution in Ecological Communities. *PLoS Biology.* 10:e1001332.

Tylianakis & Binzer. (2013). Effects of global environmental changes on parasitoidhost food webs and biological control. *Biological Control.* 75:77–86.

Tylianakis & Morris. (2017). Ecological Networks Across Environmental Gradients. *Annual Review of Ecology, Evolution, and Systematics.* 48:25–48.

Tylianakis et al. (2007). Habitat modification alters the structure of tropical hostparasitoid food webs. *Nature.* 445:202–5.

Ulanowicz. (1980). An hypothesis on the development of natural communities. *Journal of Theoretical Biology.* 85:223–45.

Ulrich. (2009). Nestedness analysis as a tool to identify ecological gradients. *Ecological Questions.* 11.

Ulrich et al. (2009). A consumer's guide to nestedness analysis. *Oikos.* 118:3–17.

Ulrich & Gotelli. (2007). Null model analysis of species nestedness patterns. *Ecology.* 88:1824–31.

Van Valen. (1973). A NEW EVOLUTIONARY LAW. *A NEW EVOLUTIONARY LAW.*

Vázquez et al. (2012). The strength of plantpollinator interactions. *Ecology.* 93:719–25.

Vellend. (2010). Conceptual synthesis in community ecology. *The Quarterly Review of Biology.* 85:183–206.

Verdú et al. (2021). Facilitation and plant phenotypic evolution. *Trends in Plant Science.* 26:913–23.

Verdú & Valiente-Banuet. (2008). The Nested Assembly of Plant Facilitation Networks Prevents Species Extinctions. *The American Naturalist.* The University of Chicago Press; 172:751–60.

Verhulst. (1938). Notice sur la loi que la population suit dans son accroissement. *Correspondance mathématique et physique.* 10:113–21.

Vermaat et al. (2009). Major dimensions in food-web structure properties. *Ecology.* 90:278–82.

Vieira & Almeida-Neto. (2015). A simple stochastic model for complex coextinctions in mutualistic networks: robustness decreases with connectance. *Ecology Letters.* 18:144–52.

Vos et al. (2009). Local Adaptation of Bacteriophages to Their Bacterial Hosts in Soil. *Science.* 325:833–3.

Watts & Strogatz. (1998). Collective dynamics of "small-world" networks. *Nature.* 393:440–2.

Weitz et al. (2005). Coevolutionary arms races between bacteria and bacteriophage. *PNAS.* National Academy of Sciences; 102:9535–40.

Weitz et al. (2013). Phage–bacteria infection networks. *Trends in Microbiology.* 21:82–91.

Whitham et al. (2003). COMMUNITY AND ECOSYSTEM GENETICS: A CONSEQUENCE

OF THE EXTENDED PHENOTYPE. *Ecology.* 84:559–73.

Whitman et al. (2005). Occurrence and growth characteristics of Escherichia coli and enterococci within the accumulated fluid of the northern pitcher plant (Sarracenia purpurea L.). *Can J Microbiol.* 51:1027–37.

Williams. (2011). Biology, Methodology or Chance? The Degree Distributions of Bipartite Ecological Networks. *PLOS ONE.* Public Library of Science; 6:e17645.

Williams & Martinez. (2000). Simple rules yield complex food webs. *Nature.* Nature Publishing Group; 404:180–3.

Wilson & Zhu. (2008). A study of graph spectra for comparing graphs and trees. *Pattern Recognition.* 41:2833–41.

Yavero lu et al. (2015). Proper Evaluation of Alignment-Free Network Comparison Methods. *Bioinformatics.*

Yeakel et al. (2014). Collapse of an ecological network in Ancient Egypt. *PNAS.* 111:14472–7.

Yodzis. (1981). The stability of real ecosystems. *Nature.* Nature Publishing Group; 289:674–6.

Yodzis & Innes. (1992). Body Size and Consumer-Resource Dynamics. *The American Naturalist.* 139:1151–75.

Yodzis & Winemiller. (1999). In search of operational trophospecies in a tropical aquatic food web. *Oikos.*:327–40.

Zeng & Wiens. (2021). Species interactions have predictable impacts on diversification. Harmon, ed. *Ecol Lett.* 24:239–48.

Zook et al. (2011). Food webs: ordering species according to body size yields high degree of intervality. *Journal of Theoretical Biology.* 271:106–13.

Annexe A

Modéliser la nature à l'aide de super ordinateurs, ça sert à quoi ?

Idaline Laigle[1], Mathilde Besson[2], Stéphanie Préfontaine[1], Madelaine Proulx[1]

[1] Département de Biologie, Université de Sherbrooke, Sherbrooke, J1K 2R1, Canada; [2] Département de Sciences Biologiques, Université de Montréal, Montréal, H2V 2J7, Canada

Contribution : ID, MB, SP et MP ont participé de façon équivalente à la rédaction de l'article. ID a édité les commentaires des éditeurs.

https://theconversation.com/modeliser-la-nature-a-laide-de-superordinateurs-ca-sert-a-quoi-92958

Contrairement à ce que l'on croit souvent, la recherche en écologie ne consiste pas seulement à observer la nature. Et certains chercheurs étudient des espèces et des écosystèmes virtuels qu'ils ont eux-mêmes créés. C'est ce qu'on appelle « faire de la modélisation ».

La modélisation en écologie consiste à reproduire des processus observés à l'aide d'équations mathématiques afin de comprendre le fonctionnement des systèmes étudiés. Ces équations représentent soit des flux, soit des variations d'abondance ou de biomasse d'espèces. Par exemple, la croissance d'une population de bactéries au cours du temps peut être représentée par une fonction exponentielle.

Les écologues peuvent également faire des prédictions sur l'état futur des écosystèmes.

Le processus de modélisation suit alors trois phases principales : l'observation, la compréhension et la prédiction. Cependant, il est parfois possible de comprendre sans pouvoir prédire, ou de prédire sans comprendre. Les avancées technologiques ont permis de développer des modèles complexes nécessitant des ordinateurs de plus en plus puissants qui permettent de lancer des simulations (prenant parfois plusieurs jours !).

A.0.1 Prévoir les effets des changements climatiques

La modélisation sert, entre autres, à comprendre les causes des changements climatiques et à prédire leur évolution. La distribution des espèces étant hautement dépendante du climat, si nous connaissons le climat futur et la distribution actuelle des espèces, nous devrions être capables de prédire une possible distribution future de celles-ci. C'est ainsi que nous pouvons imaginer la montée des espèces vers le nord.

Ces résultats sont très importants afin de préparer les gestionnaires aux changements climatiques. Ainsi, il est parfois suggéré aux forestiers de planter des essences d'arbres adaptées au climat futur. Il pourrait aussi être question de relocaliser certains vignobles, à l'image de la production de champagne désormais possible au Royaume-Uni.

Cependant, la présence d'une espèce dans un milieu ne dépend pas uniquement de

l'environnement, mais aussi de ses interactions avec d'autres espèces. Par exemple, deux espèces végétales peuvent être en compétition pour l'accès à la lumière ou l'acquisition de nutriments. Dans ce cas, l'espèce la moins compétitrice risque de se faire supplanter. Ces espèces peuvent aussi être limitées dans leur expansion par une grande pression d'herbivorie ou une absence de pollinisateurs. Ajoutées aux modèles, ces variables les rendent plus complexes.

A.0.2 Observer la dynamique des espèces

Dans la même optique, il est possible de modéliser les effets des perturbations sur des communautés entières comportant des végétaux, des herbivores et des carnivores. Une perturbation peut être progressive (comme un changement climatique) ou brutale (comme un feu ou une coupe forestière) et donc d'origine naturelle ou humaine (anthropique). Les perturbations favorisent certaines espèces (parfois des espèces exotiques et/ou pathogènes) et peuvent en éliminer d'autres.

En connaissant les interactions entre espèces, des prédictions peuvent être émises quant à l'effet d'une perturbation sur une communauté entière. Dans les modèles de réseaux trophiques (un ensemble de chaînes alimentaires) où seulement les interactions consommateurs-ressources sont prises en compte, chaque espèce consomme des ressources à des taux variables, qui sont souvent déterminés à partir de sa masse corporelle. Des espèces peuvent donc être retirées ou ajoutées aux modèles et les effets de ces modifications sur les autres espèces peuvent être calculées.

La modélisation de la dynamique de communautés d'espèces a permis de révéler des effets en cascade difficilement observables sur le terrain et difficilement prédictibles. Par exemple, la modélisation d'une grande pression de pêche sur les poissons carnivores et/ou de grande taille a montré que la pêche modifie les communautés marines dans leur ensemble, au niveau de leur composition en espèces et de leur fonctionnement. C'est à cause de cette pression de pêche que l'on a pu, par exemple, observer ces dernières années une surabondance de méduses en

Méditerranée, conséquence de la surpêche de leurs prédateurs. Cela a entraîné, en cascade, une diminution des proies des méduses telles que les petits poissons.

A.0.3   Simplifier la réalité

Les modèles peuvent être très réalistes ou très abstraits en fonction de leurs objectifs. Certaines théories écologiques ont ainsi été développées en utilisant des espèces imaginaires. Malgré cela, ces théories ont été confirmées a posteriori par des observations. La théorie de la biogéographie des îles, par exemple, permet de comprendre comment les flux d'espèces entre îles et continents façonnent les populations insulaires.
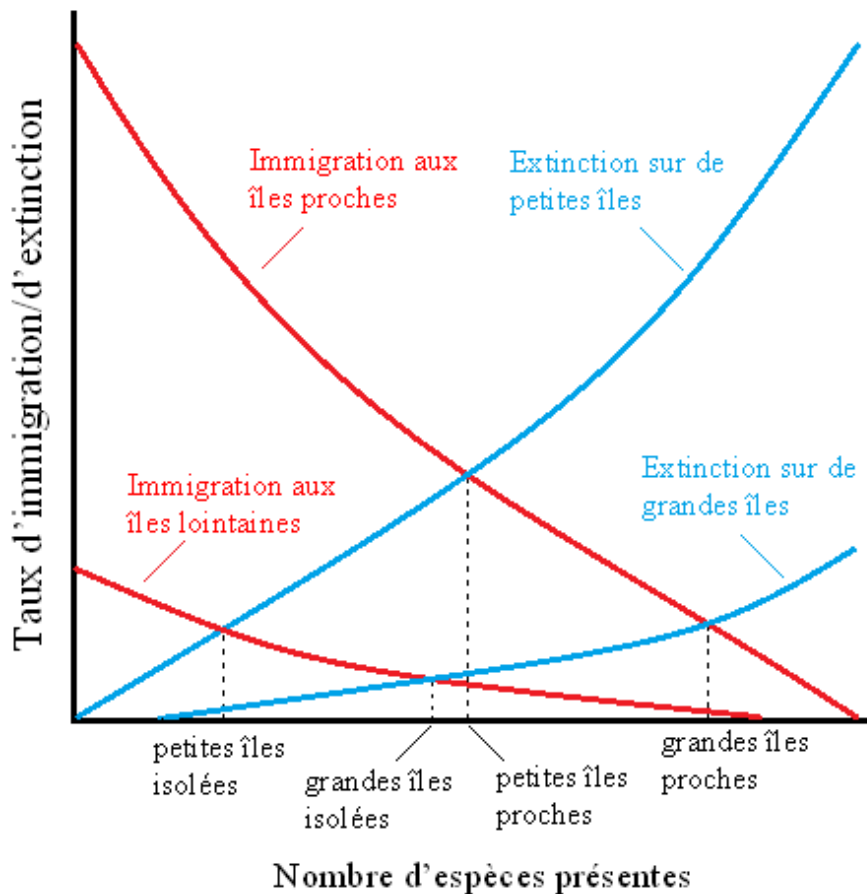

Figure A.1   Illustration de la théorie de l'ile biogéographique de R. MacArthur et E. Wilson, basée sur deux prédictions : les taux d'immigration sont plus élevés vers les îles proches d'une source de colonisateurs, et les taux d'extinction des espèces sont élevés sur les petites îles. Le modèle montre qu'il y a un plus grand nombre d'espèces dans les grandes îles proches que dans les petites îles isolées. source Wikimedia, CC BY

De nos jours, cette théorie élaborée mathématiquement est communément utilisée dans la gestion des écosystèmes. D'autres modèles sont construits et paramétrés suivant de vraies communautés. Par exemple, les communautés de micro-organismes présents dans les feuilles de forme tubulaire des Sarracénies pourpres sont utilisées pour comprendre, entre autres, comment les interactions entre espèces affectent le fonctionnement des communautés.

Il existe un gradient dans le réalisme des modèles. S'ils s'avèrent trop abstraits, cela peut induire du scepticisme quant à la justesse de leurs résultats. C'est pourquoi il est important de se rappeler qu'un modèle est construit dans le but de simplifier la réalité afin de comprendre des mécanismes précis. Un modèle comporte donc uniquement les variables d'intérêt pour une question particulière, et, comme l'a déclaré le statisticien George Box, « tous les modèles sont faux, mais certains sont utiles ».

C'est justement pour répondre à des questions diverses que de nombreux modèles, à première vue similaires, sont construits de manière un peu différente. C'est alors l'ensemble de ces modèles qui permet de comprendre différents aspects du fonctionnement des communautés et des écosystèmes, et de prédire leur réponse aux changements environnementaux.

A.0.4 Modélisation et recherche empirique

Bien que les modèles en écologie soient maintenant des outils indispensables pour comprendre le monde qui nous entoure, leur simplicité ne permet pas de saisir entièrement la complexité que l'on trouve dans la nature. Cette complexité, résultant d'un grand nombre d'espèces, d'interactions et de facteurs environnementaux, rend cependant la compréhension globale des systèmes biologiques impossible si l'on ne prend en compte que les expériences empiriques.

La modélisation reste donc indissociable des expériences de terrain. La récolte de données alimente ainsi le fonctionnement des modèles, lesquels offrent des prédictions dont la justesse peut être vérifiée empiriquement. L'utilisation conjointe en écologie de la modélisation et de la recherche empirique ouvre ainsi les portes à une nouvelle ère, à de nouvelles réponses et, bien

sûr, à de nouvelles questions.

Annexe B

Environment-host-microbial interactions shape the

Sarracenia purpurea microbiome at the continental scale

Zachary B. Freedman[1], Alicia McGrew[2,3], Benjamin Baiser[3], Mathilde Besson[4], Dominique Gravel[5], Timothée Poisot[4], Sydne Record[6], Lauren B. Trotta[2], and Nicholas J. Gotelli[7]

[1] Department of Soil Science, University of Wisconsin-Madison, Madison, WI, USA; [2] School of Natural Resources and Environment, University of Florida, Gainesville, FL, USA; [3] Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, FL, USA; [4] Département de Sciences Biologiques, Université de Montréal, Montréal, Quebec, Canada ; [5] Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, Quebec, Canada; [6] Department of Biology, Bryn Mawr College, Bryn Mawr, PA, USA; [7] Department of Biology, University of Vermont, Burlington, VT 05405;

Abstract

The importance of climate, habitat structure, and higher trophic levels on microbial diversity is only beginning to be understood. Here, we examined the influence of climate variables, plant morphology, and the abundance of aquatic invertebrates on the microbial biodiversity of the northern pitcher plant *Sarracenia purpurea*. The plant's cup-shaped leaves fill with rainwater and support a miniature, yet full-fledged, ecosystem with a diverse microbiome that decomposes captured prey and a small network of shredding and filter-feeding aquatic invertebrates that feed on microbes. We characterized pitcher microbiomes of 108 plants sampled at 36 sites from Florida to Quebec. Structural equation models revealed that annual precipitation and temperature, plant size, and midge abundance had direct effects on microbiome taxonomic and phylogenetic diversity. Climate variables also exerted indirect effects through plant size and midge abundance. Further, spatial structure and climate influenced taxonomic composition, but not phylogenetic composition. Our results suggest that direct effects of midge abundance and climate and indirect effects of climate through its effect on plant-associated factors lead to greater richness of microbial phylotypes in warmer, wetter sites.

Keywords

B.0.1 Introduction

Complex interactions between ecological communities and their environment can influence the assembly and maintenance of biodiversity across space and time. While major drivers of plant and animal distributions are well understood (Cox and Moore 2016, Rice et al. 2019), drivers of microbial distributions are less resolved (Hendershot et al. 2017). It has been established that microbial diversity often exhibits different and weaker biogeographic patterns from those of plant and animal communities (Meyer et al. 2018). Further, evidence is building that food web dynamics can exert a strong influence on microbial diversity (Koltz et al. 2018, Gralka et

190

al. 2020), that may or may not elicit functional consequences (Peschel et al. 2015, Cline et al. 2017, Beier et al. 2020). Though the extent to which these dynamics affect microbial diversity across large spatial scales is uncertain.

Phytotelmata, water-filled cavities in plants, provide an ideal system for investigating the influence of higher trophic levels on microbial diversity across large spatial scales. These unique environments are discrete units that provide similar habitats across large scales, reducing the environmental variation prevalent in studies of microbial biogeography (Sul et al. 2013, Freedman and Zak 2015). One particularly well-studied phytotelmata ecosystem is the carnivorous pitcher plant *Sarracenia purpurea*, which traps and digests insect prey in its rain-filled, pitcher-shaped leaves and is a model system in ecology (Addicott 1974, Ellison and Gotelli 2003, Gotelli and Ellison 2006). To decompose prey, *S. purpurea* relies on a food web that resides in the pitcher composed of microorganisms, protozoa, rotifers, and dipteran larvae, among other organisms. This micro-ecosystem has been used to study keystone predation, community assembly, tipping points, and evolution, among other community dynamics (*e.g.*, Addicott 1974, terHorst et al. 2010, Miller 2012, Sirota et al. 2013). Furthermore, *S. purpurea* has a wide distribution across North America (Fig. B.1, which provides the opportunity to explore the biogeography of an entire food web (Buckley et al. 2010, Baiser et al. 2012, Gray et al. 2012).

Although the microbial community is responsible for the main ecosystem function, decomposition, in the *S. purpurea* system (Butler et al. 2008), current understanding of the degree to which climatic-, pitcher-habitat-, and food-web-associated factors influence the *S. purpurea* microbiome across continental spatial scales is nascent. Macroecological theory has different predictions regarding the relative importance of direct and indirect drivers of microbial diversity. For example, the metabolic theory of ecology (Brown et al. 2004) predicts that climatic conditions should directly constrain microbial diversity as warmer environments are expected to support greater turnover due to increasing metabolic demand, regardless of local species composition (Fig. B.2); Gray et al. 2016). This theory has been used as support for
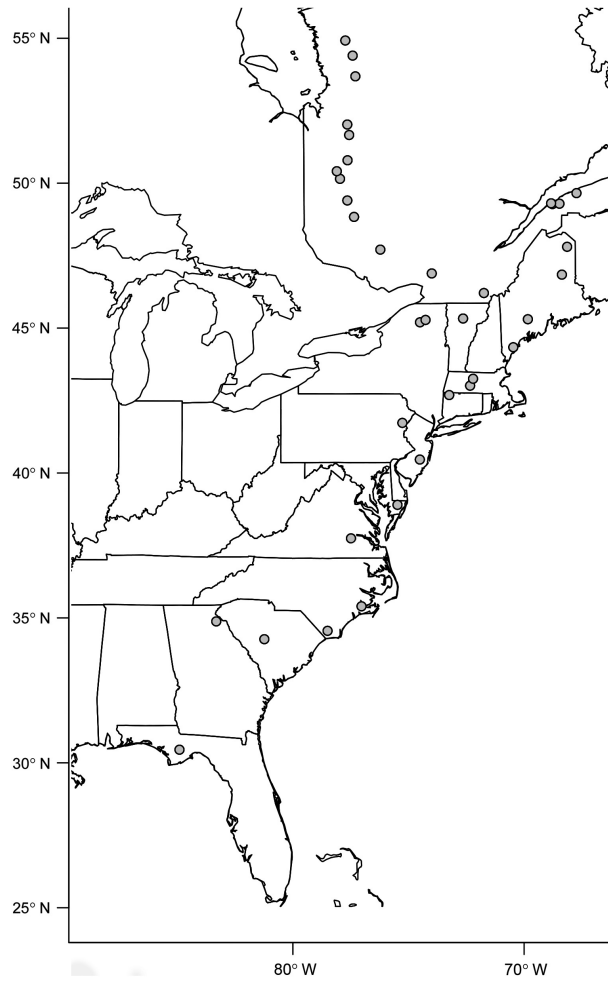
Figure B.1 Map detailing the location of the 36 sampling sites in this study. Circles represent sampling sites. Three microbial samples were analyzed from each location. One site (in the southeastern coastal plain of Georgia, USA) was omitted from the map by landowner request.

macroecological patterns like the latitudinal diversity gradient, where species richness generally increases from high to low latitudes (Pianka 1966). Conversely, the temperature-dependent consumer-resource theory (Vasseur and McCann 2005) predicts an increased top-town influence of predators on producers in warmer environments. Indeed, the presence of a keystone predator (*Wyeomyia smithii*) can increase bacterial richness in *S. purpurea* pitchers (Peterson et al. 2008), and this response may be dependent on climate as bacterial thermal tolerances can differ from that of bacterivorous consumers (Parain et al. 2016). As a result, this theory would support an indirect impact of climate through food web-associated factors on microbial diversity (Fig. B.2 B). Alternatively, the Habitat Amount Hypothesis (Fahrig 2013) predicts

that microbial diversity will increase with the amount of habitat type (e.g., pitcher size and morphology). As warmer, wetter climates select for larger pitchers (Ellison et al. 2004), and variation in the pitcher habitat can influence members of its captive ecosystem (Gotelli and Ellison 2006), this hypothesis would predict an indirect effect of climate through plant-associated factors influencing microbial diversity (Fig. B.2 C). Last, niche theory (Hutchinson 1957) predicts a sequential turnover of species along the climatic gradient that can be attributed to multidimensional changes in environmental conditions and resources. This would result in climatic-, plant-, and food-web-associated factors together influencing the *S. purpurea* microbiome (Fig. B.2 D).

In this study, we investigated the influence of climatic-, pitcher-habitat-, and food-web-associated factors in shaping the bacterial communities inhabiting *S. purpurea* pitchers along its North American latitudinal range. Specifically, we draw on macroecological theory to develop and test four hypotheses (Fig. B.2) where microbial diversity is directly affected by climate (H1; Fig. B.2 A), or indirectly affected by climate's interactions with characteristics of the food web (H2; Fig. B.2 B), pitcher-habitat-associated factors (*e.g.*, plant morphology, water volume, and pH; H3; Fig. B.2 C), or both pitcher-habitat- and food-web-associated factors (H4; Fig. B.2 D). We tested these hypotheses with a survey of microbial assemblages collected from 108 pitchers from 36 sites across 23.53º of latitude, from Florida, USA to northern Quebec, Canada.

B.0.2 Material and Methods

B.0.3 Site selection and sample processing

We collected fluid from 108 pitchers at 36 sites across *S. purpurea*'s North American latitudinal range (30.197º N-53.722º N; Appendix S1: Section S1; Fig. B.1). Selected sites had a minimum of seven individuals and maximized latitudinal coverage of *S. purpurea*'s geographic distribution. Sampling of each site was timed to ensure that plants were sampled between four

and six weeks after new pitchers opened, which is sufficient time for pitchers to capture prey and be colonized by their aquatic food web (Buckley et al. 2010).

At each site, we marked plants at 5-m intervals along a 120-m transect and selected three pitchers randomly for further analysis. For each selected plant, we measured the diameter of each plant, pitcher liquid was collected from the largest new pitcher, and the leaf was collected for morphometry measurements (Ellison and Gotelli 2003). We then homogenized the pitcher fluid, transferred it to a sterile 50-mL Falcon tube and immediately recorded water volume and pH. Captured prey and pitcher invertebrates (flesh flies, midges, and mosquitos) were visually counted. Samples were stored in the field at $-20°C$ in an electric chest freezer. Upon return to the lab, we stored subsamples of liquid at $-20°C$ and at $-80°C$.

B.0.4 Microbiome analysis

We extracted total DNA from 300 μL of homogenized pitcher fluid and the V4 region of the 16S rRNA gene was amplified and sequenced (Appendix S1: Section S1). Bacterial phylotypes were delineated by clustering operational taxonomic units (OTUs) at the $\pm$ 97% sequence similarity in QIIME (Version 1.9.1; Caporaso et al. 2010), from which α and β taxonomic and phylogenetic diversity were determined. We rarefied the data set to 10,000 sequences per sample prior to downstream α - and β-diversity analyses.

B.0.5 Statistical analysis

We employed linear mixed models with site as a random effect to test for associations between bacterial taxonomic (Chao1) and phylogenetic (MPD, SES.MPD) α-diversity and latitude. Briefly, Chao1 is a nonparametric richness estimator, MPD is the mean phylogenetic distance among all pairs of phylotypes within a community, and SES.MPD is the standardized effect size of MPD compared to a null model (Appendix S1: Section S1). We then performed structural equation modeling (SEM) to test hypotheses of how climate influences pitcher-habitat-

and food-web-associated factors as well as their collective influence on the taxonomic and phylogenetic α-diversity of the *S. purpurea* microbiome (*Fig 2*; Appendix S1: Section S1). We assessed each model's goodness-of-fit using tests of directed separation (Shipley 2000). Partial redundancy analysis (RDA) and variance partitioning were then conducted to test the role of climatic-, pitcher-habitat-, and food-web-associated factors on bacterial taxonomic and phylogenetic composition (Appendix S1: Sections S2 and S3). Partial RDAs and distance-based RDAs (dbRDA) were conditioned on distance-based Moran's Eigenvector Maps (dbMEM) to partition out the role of spatial structure.

B.0.6 Results

B.0.7 Diversity of the S. purpurea microbiome across its North American latitudinal range

In total, 16,157,518 high-quality sequences were obtained from 108 pitchers across the 36 sites. After rarefaction, sites averaged 336-1,141 OTUs, and there were 22,307 total OTUs across all sites (Good's coverage of 0.986 $\pm$ 0.003 [mean $\pm$ SE]). The most abundant OTUs at the individual pitcher level were also the most widely distributed across the latitudinal range, as there was a positive correlation between OTU abundance and number of pitchers occupied (F1, 105, 029 = 2.2 $\times$ 105; R^2 = 0.68, P < 0.01; Appendix S1: Section S4 and Fig. S4). Across all sites, S. purpurea pitcher microbiomes were dominated by members of the phyla Proteobacteria (54 - 79% of community; Appendix S1: Table S3), Bacteroidetes (6 - 21%), and Actinobacteria (4 - 20%), while members of the Firmicutes (1 - 7%) and Acidobacteria (1 - 2%) were less abundant. Among the proteobacterial phylum, the most dominant taxa were members of the class beta-proteobacteria (14 - 39%), alpha-proteobacteria (12 - 35%), and gamma-proteobacteria (11 - 39%).

S. purpurea microbiome taxonomic -diversity (*i.e.*, Chao1 diversity) decreased with latitude toward the north (coefficient = −17.77; t1,34 = −3.30; P < 0.005; R^2 = 0.17; Fig. B.3).

Phylogenetic α-diversity measured as MPD showed no relationship with latitude (P = 0.16) but standardized MPD (SES.MPD) increased slightly with latitude (coefficient = 0.10; t1,34 = 2.18; P < 0.05; R^2 = 0.06).

B.0.8   Drivers of the S. purpurea microbiome biogeography

While SEMs tested all hypothesized relationships (Fig. B.2), results and subsequent discussion will focus on variables that directly or indirectly influenced microbial diversity. Full SEM results may be found in Appendix S1: Section S5. Climate and food web variables directly influenced bacterial taxonomic α-diversity (*i.e.,* Chao1 diversity). Chao1 increased in warmer, wetter climates (Fig. B.3; standardized coefficient (std. coef.) = 0.32; P = 0.02) and increased with midge abundance (std. coef. = 0.17; P = 0.04). The direct effect of climate was almost twice as large the direct effect of midges. Climate also indirectly influenced Chao1 through its effect on pitcher morphology and water volume (Fig. B.3 C and E). Warmer, rainier climates resulted in greater pitcher water volumes directly (std. coef. = 0.32, P < 0.005) and indirectly through plant morphology (std. coef. = 0.16; P values unavailable for indirect effects, which are the product of the standardized coefficients along the path; Appendix S1: Section S1). Pitchers with larger volumes of water harbored greater abundances of midges (unstandardized coef. = 1.14, P < 0.001; standardized coefficients unavailable for non-Gaussian error distributions) and, as noted above, midges increased bacterial taxonomic α-diversity. Overall, 26% of the variation in the *S. purpurea* microbiome taxonomic α-diversity was described by the SEM model (Fisher's C = 20.10, df = 14, P = 0.13).

Both measures of phylogenetic α-diversity (*i.e.,* MPD and SES.MPD) responded similarly (*i.e.,* same significant paths and signs of coefficients) to climatic-, pitcher-habitat-, and food-web-associated variables. For brevity, only MPD results will be presented (SES.MPD results can be found in Appendix S1: Section S5 and Table S7). *Sarracenia purpurea* microbiome phylogenetic α-diversity was directly influenced by climatic-, pitcher-habitat-, and food-web-

associated factors. MPD decreased in warmer, rainier sites (Fig. B.3 D and F; std. coef. = −0.37; P < 0.005) and increased in larger pitchers (std. coef. = 0.36; P < 0.005) and pitchers with greater midge abundances (std. coef. = 0.23; P < 0.05). Climate indirectly influenced MPD along the same paths as it did for Chao1 diversity (*i.e.*, through volume and midge abundance), and it also indirectly affected MPD through plant morphology. Note that the direct effect of climate on MPD is negative, while the indirect effects of climate on MPD are positive. Overall, 21% of the variation in MPD was explained by our model (Fisher's C = 16.01, df = 14, P = 0.31). Coefficient tables for all SEM equations can be found in Appendix S1: Tables S5-S7.

The constraints in partial RDA model described only 2% of the variation in taxonomic composition (Appendix S1: Section S6) with water volume (P = 0.05) showing significance. Variance partitioning indicated that the food web, plant morphology, climate, and spatial structure (*i.e.*, dbMEMs) together accounted for 14% of the variation in community composition. The variance partitions for spatial structure (5%), shared spatial structure and climate (4%) explained most of the variation. For phylogenetic composition, the constraints and conditions (*i.e.*, dbMEMs) in the partial dbRDA described < 1% of the variation.

B.0.9   Discussion

Bacterial communities inhabiting *S. purpurea* pitchers exhibit biogeographic structure across the plant's North American range, driven by the direct and indirect effect of climate through pitcher-habitat- and food-web-associated factors. The empirical data from this study supports hypothesis H4: the *S. purpurea* microbiome is directly affected by climate and indirectly affected by climate through interactions with both pitcher-habitat- and food-web-associated factors (*Figs. 2* and *3*). In this way, our findings are also consistent with the latitudinal diversity gradient and provide support for the metabolic theory of ecology at the continental scale. Conversely, there was no support for the temperature-dependent consumer-resource theory, as

the detritus-shredding midge, but not mosquito larvae (*i.e.*, a keystone predator), emerged as the food web-associated factor most influential to pitcher microbiome diversity. Taxonomic α-diversity was greatest in pitchers that experienced warm temperatures and high levels of precipitation, generally located at low latitudes, and it decreased with increasing latitude where sites are relatively cooler and drier. The inverse relationship between taxonomic diversity and latitude is consistent with observations of marine and aquatic bacterial diversity (Fuhrman et al. 2008, Ladau et al. 2013) and diversity for many other organisms (*i.e.*, the latitudinal diversity gradient; Fischer 1960). Conversely, bacterial communities exhibited a weak negative latitudinal-phylogenetic α-diversity relationship (*i.e.*, greater phylogenetic clustering at lower latitudes), a trend that has been previously observed with bacteria (Andam et al. 2016). Given that higher temperatures are associated with high rates of speciation (Allen et al. 2006), it is plausible that in southern *S. purpurea* pitcher ecosystems the high taxonomic and low phylogenetic α-diversity may be maintained by higher diversification rates that generate rich assemblages of closely related phylotypes as compared to northern *S. purpurea* ecosystems.

The relationship between climate and *S. purpurea* morphology has consequences for pitcher bacterial diversity. For example, wetter climates generally have larger pitchers (Ellison and Gotelli 2003) that can hold a greater volume of rainwater, which in turn can influence members of the *S. purpurea* food web (Gotelli and Ellison 2006). Here, our results show that climatic-driven changes in pitcher morphology increase phylogenetic α-diversity of pitcher microbial communities. Larger pitchers potentially provide more habitat heterogeneity (*i.e.*, sub-habitats) and total potential energy (*i.e.*, prey), which together can support distinct microbial communities (Krieger and Kourtev 2012). The observation that the *S. purpurea* microbiome is partially constrained by pitcher-associated factors has been previously noted (Peterson et al. 2008) and is not surprising, as the *S. purpurea* microbiome has had the opportunity to coevolve and adapt since Sarracenia diverged from its congeners $\sim$ 23 million years ago (Ellison et al. 2012).

Pitcher plant morphology also interacts with members of higher trophic levels in the food web, particularly the shredding midge *Metriocnemus knabi*, to influence bacterial diversity. Midge abundance has a positive relationship with pitcher size (Buckley et al. 2010), and our results suggest that this pitcher-morphology-midge relationship leads to greater bacterial taxonomic and phylogenetic α-diversity. *Metriocnemus knabi* is the only shredder in the system and is part of a commensal "processing chain" where it shreds detritus into smaller pieces, thereby facilitating bacterial growth by increasing surface area (Heard 1994) and increasing decomposition rates of detritus (Butler et al. 2008, Baiser et al. 2011). Increased bacterial density and smaller detrital particle sizes can increase the growth rate of the mosquito, *W. smithii*, the top-level consumer in the food web (Heard 1994). Whereas it has been proposed that *W. smithii* is a keystone predator that constrains *S. purpurea* microbiome diversity at the local scale (Peterson et al. 2008), results presented here suggest that the shredding midge is more influential in driving taxonomic and phylogenetic diversity of the *S. purpurea* microbiome at the continental scale. While *M. knabi* can facilitate bacterial diversity by providing resources through detrital shredding (Heard 1994, Butler et al. 2008) we did not assess the role of mosquito instar and thus could not directly test the role of the full "processing chain" described by Heard (1994).

The relatively similar conditions that the *S. purpurea* microbiome experiences (*e.g.*, same plant species and food web) across a large spatial scale provide an ideal standardized system to sample the microbiome and explore external sources of variation in composition. Taxonomic composition was mainly driven by spatial factors and climate, indicating that microbiomes that are geographically closer to one another and are more alike in climatic conditions have similar microbial composition. On the other hand, phylogenetic composition was not driven by any of the variables in our study. This suggests that while dispersal limitation and filtering due to climate may influence taxonomic composition, most microbial clades are distributed across the range of climate, plant habitat conditions, and food web composition in our study.

Overall, we show that climatic conditions shape *S. purpurea* microbiome diversity at the continental scale both directly and indirectly through their effect on local variables related to the host pitcher habitat and members of higher trophic levels. In this study, we considered food-web-associated influences on *S. purpurea* bacterial diversity by quantifying the abundance of higher trophic-level organisms. To advance the study of microbial diversity in a food web context, future studies should aim to more closely tie trophic dynamics with microbial diversity through methods such as stable isotope tracing. Altogether, considering the direct and indirect effects of climate, food web structure and microbiome host conditions across broad spatial scales can lead to greater insights into the macroecology of microbes.
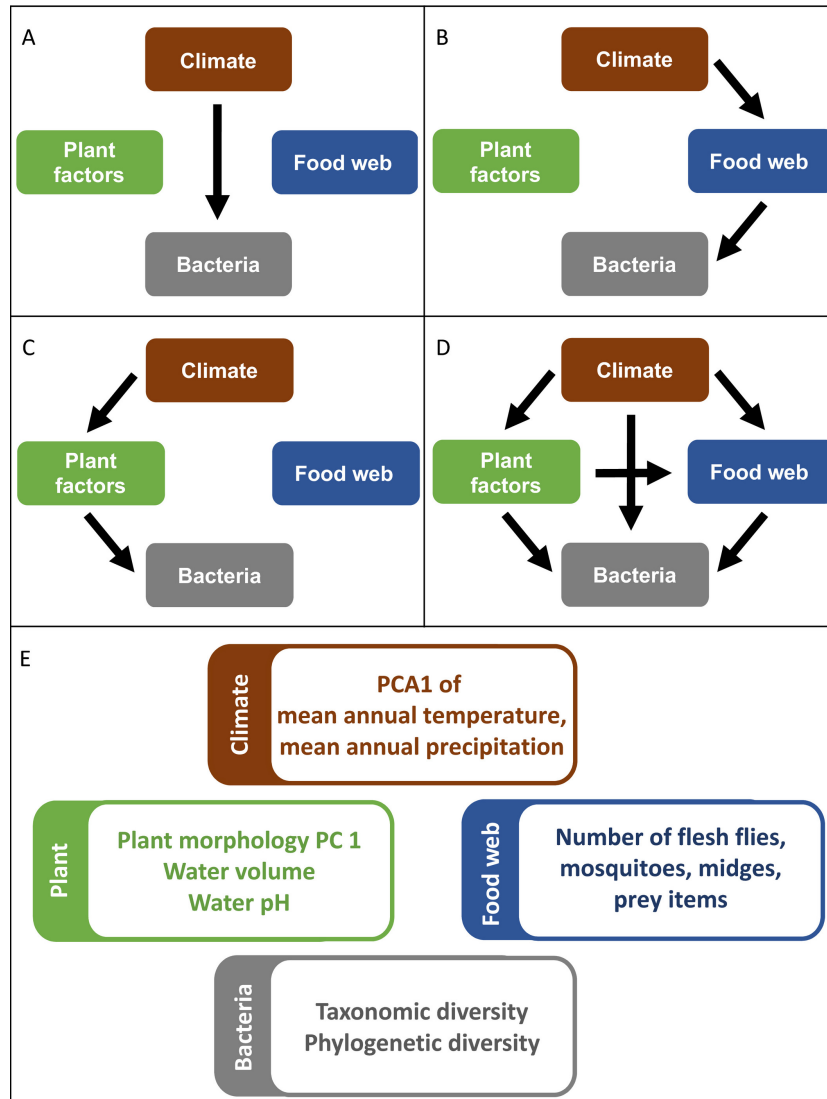
Figure B.2 (A) Direct and (B-D) indirect hypotheses by which climate impacts the bacterial diversity within *Sarracenia purpurea* pitchers. (E) The climatic-, pitcher-habitat-, and food-web-associated factors measured in this study. Plant morphology metrics included pitcher length and width, keel width, mouth diameter, and lip width.
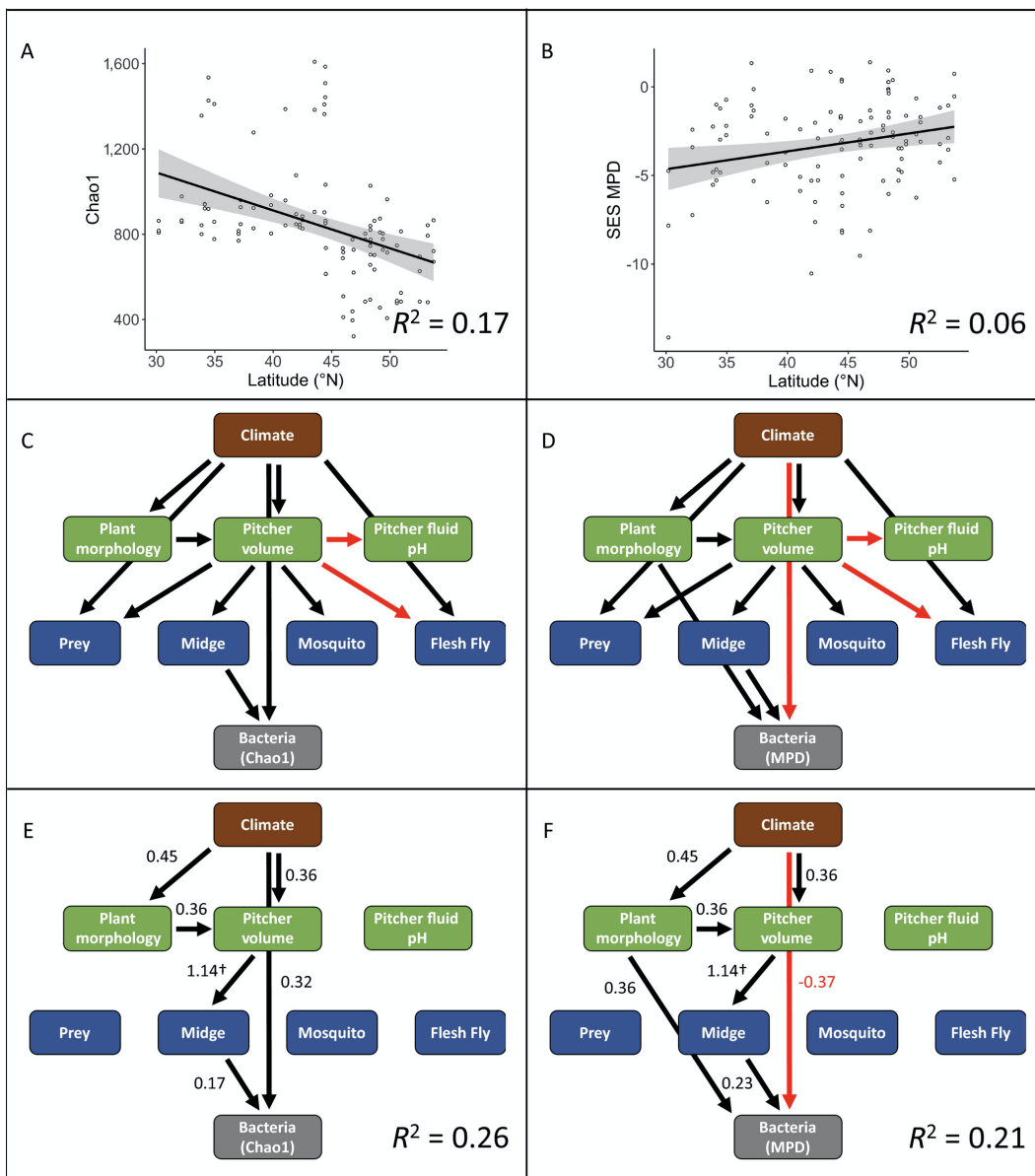
Figure B.3 (A, B) Biplots of diversity vs. latitude and (C-F) results of structural equation models (SEM) linking climate-, pitcher-habitat-, and food-web-associated factors to bacterial taxonomic (C, E) and phylogenetic (D, F) diversity. All significant paths are shown in panels C and D whereas only significant paths leading to bacterial Chao1 (model R2 = 0.26) and mean phylogenetic distance (MPD; model R2 = 0.21) are shown in panels E and F. Black and red arrows denote positive and negative coefficients, respectively. Standardized coefficients are presented unless marked by a dagger (†).

B.0.10 Acknowledgements

B.0.11 Supporting information

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/ecy.3308/suppinfo

B.0.12 Data availability

Raw sequence reads for the pitcher plant microbiome are available from the National Center for Biotechnology Information(NCBI) under accession no. PRJNA641293.

B.0.13 References

Addicott, J. F. 1974. Predation and prey community structure: an experimental study of the effect of mosquito larvae on theprotozoan communities of pitcher plants. Ecology 55:475-492.

Allen, A. P., J. F. Gillooly, V. M. Savage, and J. H. Brown. 2006. Kinetic effects of temperature on rates of genetic divergenceand speciation. Proceedings of the National Academy ofSciences USA 103:9130-9135.

Andam, C. P., J. R. Doroghazi, A. N. Campbell, P. J. Kelly, M.J. Choudoir, and D. H. Buckley. 2016. A latitudinal diversitygradient in terrestrial bacteria of the genusStreptomyces.mBio 7:e02200-02215.

Baiser, B., R. S. Ardeshiri, and A. M. Ellison. 2011. Speciesrichness and trophic diversity increase decomposition in a co-evolved food web. PLoS ONE 6:e20672.

Baiser, B., N. J. Gotelli, H. L. Buckley, T. E. Miller, and A. M.Ellison. 2012. Geographic variation in network structure of anearctic aquatic food web. Global Ecology and Biogeography 21:579-591.

Beier, S., A. F. Andersson, P. E. Galand, C. Hochart, J. B.Logue, K. McMahon, and S. Bertilsson. 2020. The environ-ment drives microbial trait variability in aquatic habitats.Molecular Ecology 29:4605-4617.

Brown, J. H., J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B.West. 2004. Toward a metabolic theory of ecology. Ecology 85:1771-1789.

Buckley, H. L., T. E. Miller, A. M. Ellison, and N. J. Gotelli.2010. Local to continental-scale variation in the richness andcomposition of an aquatic food web. Global Ecology andBiogeography 19:711-723.

Butler, J. L., N. J. Gotelli, and A. M. Ellison. 2008. Linking thebrown and green: nutrient transformation and fate in the Sarraceniamicroecosystem. Ecology 89:898-904.

Caporaso, J. G. et al. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7(5):335-336.

Cline, L. C., D. R. Zak, R. A. Upchurch, Z. B. Freedman, and A. R. Peschel. 2017. Soil microbial communities and elk foraging intensity: implications for soil biogeochemical cyclingin the sagebrush steppe. Ecology Letters 20:202-211.

Cox, C. B., and P. D. Moore. 2016. Biogeography, an ecologicaland evolutionary approach. Wiley, New York, New York,USA. https://www.wiley.com/en-us/Biogeography

Ellison, A. M., H. L. Buckley, T. E. Miller, and N. J. Gotelli. 2004. Morphological variation inSarracenia purpurea(Sar-raceniaceae): geographic, environmental, and taxonomic correlates. American Journal of Botany 91:1930-1935.

Ellison, A. M., E. D. Butler, E. J. Hicks, R. F. C. Naczi, P. J.Calie, C. D. Bell, and C. C. Davis. 2012. Phylogeny and bio-geography of the carnivorous plant family Sarraceniaceae. PLoS ONE 7:e39291.

Ellison, A. M., and N. J. Gotelli. 2003. Evolutionary ecology of carnivorous plants. Trends in Ecology and Evolution 16:623-629.

Fahrig, L. 2013. Rethinking patch size and isolation effects: the habitat amount hypothesis. Journal of Biogeography 40:1649-1663.

Fischer, A. G. 1960. Latitudinal variations in organic diversity. Evolution 14:64-81.

Freedman, Z., and D. R. Zak. 2015. Soil bacterial communities are shaped by temporal and environmental filtering: evidencefrom a long-term chronosequence. Environmental Microbiology 17:3208-3218.

Fuhrman, J. A., J. A. Steele, I. Hewson, M. S. Schwalbach,M. V. Brown, J. L. Green, and J. H. Brown. 2008. A latitudinal diversity gradient in planktonic marine bacteria. Pro-ceedings of the National Academy of Sciences USA105:7774-7778.

Gotelli, N. J., and A. M. Ellison. 2006. Food-web models predict species abundances in response to habitat change. PLoSBiology 4:1869-1873.

Gralka, M., R. Szabo, R. Stocker, and O. X. Cordero. 2020. Trophic interactions and the drivers of microbial communityassembly. Current Biology 30:R1176-R1188.

Gray, S. M., D. M. Akob, S. J. Green, and J. E. Kostka. 2012. The bacterial composition within the Sarracenia purpurea model system: local scale differences and the relationship with the other members of the food web. PLoS ONE 7:e50969.

Gray, S. M., T. Poisot, E. Harvey, N. Mouquet, T. E. Miller, and D. Gravel. 2016. Temperature and trophic structure are driving microbial productivity along a biogeographical gradient. Ecography 39:981-989.

Heard, S. B. 1994. Pitcher-plant midges and mosquitoes: a processing chain commensalism. Ecology 75:1647-1660.

Hendershot, J. N., Q. D. Read, J. A. Henning, N. J. Sanders,and A. T. Classen. 2017. Consistently inconsistent drivers of microbial diversity and abundance at macroecological scales. Ecology 98:1757-1763.

Hutchinson, G. E. 1957. Concluding remarks. Cold Spring Harbor Symposium on Quantitative Biology 22:415-427.

Koltz, A. M., A. T. Classen, and J. P. Wright. 2018. Warmingreverses top-down effects of predators on belowground ecosystem function in Arctic tundra. Proceedings of the National Academy of Sciences USA 115:E7541-E7549.

Krieger, J. R., and P. S. Kourtev. 2012. Bacterial diversity in three distinct sub-habitats within the pitchers of the northern pitcher plant, Sarracenia purpurea. FEMS Microbiology Ecology 79:555-567.

Ladau, J., T. J. Sharpton, M. M. Finucane, G. Jospin, S. W.Kembel, J. O'Dwyer, A. F. Koeppel, J. L. Green, and K. S.Pollard. 2013. Global marine bacterial diversity peaks at high latitudes in winter. ISME Journal 7:1669-1677.

Meyer, K. M., H. Memiaghe, L. Korte, D. Kenfack, A. Alonso, and B. J. M. Bohannan. 2018. Why do microbes exhibit weak biogeographic patterns? ISME Journal 12:1404-1413.

Miller, T. E. 2012. Testing successional hypotheses of stability, heterogeneity, and diversity in pitcher-plant inquiline communities. Oecologia 170:243-251.

Parain, E. C., D. Gravel, R. P. Rohr, L. F. Bersier, and S. M.Gray. 2016. Mismatch in microbial food webs: predators but not prey perform better in their local biotic and abiotic conditions. Ecology and Evolution 6:4885-4897.

Peschel, A. R., D. R. Zak, L. C. Cline, and Z. Freedman. 2015. Elk, sagebrush, and saprotrophs: Indirect top-down control on microbial community composition and function. Ecology 96:2383-2393.

Peterson, C. N., S. Day, B. E. Wolfe, A. M. Ellison, R. Kolter, and A. Pringle. 2008. A keystone predator controls bacterial diversity in the pitcher-plant (Sarracenia purpurea) microecosystem. Environmental Microbiology 10:2257-2266.

Pianka, E. R. 1966. Latitudinal gradients in species diversity: a review of concepts. American Naturalist 100:33-46.

Rice, A., P. Smarda, M. Novosolov, M. Drori, L. Glick, N.Sabath, S. Meiri, J. Belmaker, and I. Mayrose. 2019. The global biogeography of polyploid plants. Nature Ecology & Evolution 3:265-273.

Shipley, B. 2000. A new inferential test for path models basedon directed acyclic graphs. Structural Equation Modeling 7:206-218.

Sirota, J., B. Baiser, N. J. Gotelli, and A. M. Ellison. 2013. Organic-matter loading determines regime shifts and alternative states in an aquatic ecosystem. Proceedings of the National Academy of Sciences USA 110:7742-7747.

Sul, W. J., T. A. Oliver, H. W. Ducklow, L. A. Amaral-Zettler, and M. L. Sogin. 2013. Marine bacteria exhibit a bipolar distribution. Proceedings of the National Academy of Sciences USA 110:2342-2347.

terHorst, C. P., T. E. Miller, and D. R. Levitan. 2010. Evolution of prey in ecological time reduces the effect size of predators in experimental microcosms. Ecology 91:629-636.

Vasseur, D. A., and K. S. McCann. 2005. A mechanistic approach for modeling temperature-dependent consumer-resource dynamics. American Naturalist 166:184-198.