Université de Montréal

High Throughput DNA barcoding to assess the diversity of Laurentian insects

Par

Malek Kalboussi

Département de sciences biologiques

Faculté des arts et des sciences

Mémoire présenté

En vue de l'obtention du grade de maîtrisât en sciences biologiques

Août, 2022

Université de Montréal

Département des sciences biologiques, Faculté des arts et des sciences

Ce mémoire intitulé

High Throughput DNA barcoding to assess the diversity of Laurentian insects

Présenté par

Malek Kalboussi

A évalué par un jury composé des personnes suivantes

Étienne Léveillé-Bourret

Membre du jury

Simon Joly

Membre du jury

Colin Favret

Directeur de recherche

Résumé

La grande diversité d'insectes et la quantité de spécimens recueillis lors de l'échantillonnage constituent les plus grands défis de la systématique des insectes. Le tri des échantillons au niveau des espèces est nécessaire avant qu'ils puissent être utilisés pour des enquêtes sur les modèles de biodiversité. En raison de l'obstacle taxonomique, le manque d'expertise taxonomique, de nombreuses études sur la diversité des insectes classe les spécimens en Unités Morphologiques Opérationnelles (MorphOTUs), aussi appelées morpho-espèces, en désignant les groupes définis subjectivement en fonction de caractéristiques morphologiques évidentes. Cependant, il est long et douteux de définir avec précision les limites des espèces en se fondant sur les MorphOTUs, surtout dans les groupes où il y a de minuscules insectes et une grande similarité au niveau des espèces, comme chez les Hyménoptères. Le codage à barres de l'ADN, une approche taxonomique discriminatoire qui utilise des séquences d'ADN, a accéléré la classification taxonomique et peut être une approche alternative aux MorphOTUs. Cependant, il est crucial d'utiliser une stratégie fiable et économique de codage à barres ADN pour traiter un grand nombre d'échantillons. En outre, le codage à barres d'ADN devrait fonctionner avec les espèces problématiques dans l'entomologie moléculaire comme on l'observe parfois avec les hyménoptères.

Afin de mettre en œuvre une évaluation rapide de la biodiversité des Hyménoptères, optimiser les de étapes de barcodage d`ADN (extraction d'ADN, amplification par PCR et séquençage) était le premier objectif de ce projet de recherche. On a testé et optimisé une extraction d'ADN arrivant à une méthode coûtante 0,20 dollars par spécimen. On a validé la performance adéquate des mini-codes à barres d'ADN, réduits en taille à 313bp, pour établir une classification d'Unités Taxonomiques Opérationnelles Moléculaires (MOTUs) comparable à celle du codage à barres d'ADN couramment utilisé, de longueur de 658bp. On a adopté ce protocole optimisé pour le codage à barres de 517 spécimens d'Hyménoptères échantillonnés par des pièges aspirateurs situés dans la forêt laurentienne de l'Est du Canada. Avec le séquençage multiplexé à haut débit Illumina, impliquant des amplicons étiquetés, on a obtenu des mini-codes à barres pour 88% des spécimens. Le coût et le temps nécessaires pour générer des données MOTU, grâce à notre approche de codage à barres d'ADN, étaient environ la moitié de celui de la classification morphologique en MorphOTUs.

Le deuxième objectif de ma recherche était de comparer l'efficacité du tri morphologique des MorphOTUs avec l'identification moléculaire et la délimitation par MOTUs. On a démontré

une forte congruence entre l'identification morphologique et moléculaire au niveau taxonomique de la famille dans la base de données Barcode of Life (BOLD) et GenBank (93 %), alors que seulement 18 % des mini-codes à barres ont été attribués à des identifications plus précises (genre ou espèce). La délimitation moléculaire s'est faite avec quatre méthodes de regroupement différentes (basée sur la distance : Découverte automatique de l'écart de codes à barres (ABGD) et Assemblage des espèces par partitionnement automatique (ASAP) ; basée sur un dendrogramme : Coalescente mixte généralisée du yule (GMYC) et Processus bayésien de l'arbre de poisson (bPTP)). En générale, les méthodes moléculaires ont plus que doublé la diversité estimée des MorphOTUs des Hyménoptères. Les MOTUs étaient en grande partie incompatibles avec les MorphOTUs (ratio d'appariement <0,35). Les méthodes basées sur la distance ont donné des résultats plus conformes au tri morphologique que les méthodes basées sur les arbres, en particulier dans la superfamille des Chalcidoidea.

Compte tenu de la comparaison entre le coût et le temps des méthodes de classification moléculaire et morphologique, nos résultats suggèrent que le codage à barres mini-ADN pour estimer la diversité des espèces d'Hyménoptères est plus économique que le tri par MorphOTU. Cependant, bien que les méthodes MorphOTU et MOTU aient donné de nombres unités taxonomiques différentes, les analyses de la diversité utilisées actuellement tiennent compte de l'abondance et d'autres paramètres. On n'a pas évalué si les MorphOTUs et les protocoles d'entente donneraient des résultats suffisamment équivalents dans la recherche réelle sur les diversités α et β, c'est-à-dire pour évaluer s'ils pouvaient tout de même tous deux être utiles.

Mots-clés : Codage à barres d'ADN haut débit, Hyménoptères, biodiversité, délimitation des espèces, COI.

Abstract

The great insect diversity and the quantity of insect specimens collected during sampling constitute the biggest challenges facing insect systematics. Sorting samples to the species level is necessary before they can be used for investigations of biodiversity patterns. Because of the Taxonomic impediment, the lack of taxonomic expertise, many insect diversity studies sort specimens to Morphological Operational Taxonomic Units (MorphOTUs), also known as morphospecies, classifying subjectively defined groups based on obvious morphological features. However, accurately defining species boundaries based on MorphOTUs is time consuming and questionable, especially in groups with tiny insects and great species-level similarity such as Hymenoptera. DNA barcoding, a taxonomic discriminatory approach that employs DNA sequences, has accelerated taxonomic classification and may be an alternative approach to MorphOTUs. However, it is crucial to use a reliable and economic DNA barcoding strategy to deal with a large number of samples. Additionally, DNA barcoding should work with species problematic in molecular entomology as is sometimes observed with Hymenoptera.

In order to implement a rapid biodiversity assessment of Hymenoptera, optimizing the DNA barcoding steps (DNA extraction, PCR amplification, and DNA sequencing) was the first objective of this present research. We tested and optimized a DNA extraction arriving at a method costing 0.20CAD per specimen. We validated the adequate performance of 313bp mini-barcodes for establishing Molecular Operational Taxonomic Units (MOTUs) classification, comparable to that of the commonly used full-length DNA barcode of 658bp. We adopted this optimized protocol to barcode 517 Hymenoptera specimens sampled with suction traps located in the Laurentian Forest of eastern Canada. With multiplexed Illumina high throughput sequencing of tagged amplicons, we obtained mini-barcodes for 88% of specimens. The cost and time taken to generate MOTU data through our DNA barcoding approach was approximately twice that of morphological identification for MorphOTU designation.

The second objective of my research was to compare the efficacy of morphological sorting of MorphOTUs with the molecular identification and delimitation of MOTUs. We found a high taxonomic congruence between morphological and molecular identification at family level in Barcode of Life (BOLD) and GenBank databases (93%), whereas only 18% of mini-barcode data was assigned to more precise identification (genus or species). Molecular delimitation based on four different clustering methods (distance-based: Automatic Barcode Gap Discovery

(ABGD) and Assemble Species by Automatic Partitioning (ASAP); tree-based: Generalized Mixed Yule Coalescent (GMYC) and Bayesian Poisson Tree Processes (bPTP)) resulted in more than doubling the estimated diversity of Hymenoptera as compared to MorphOTUs. The MOTUs were largely incongruent with MorphOTUs (match ratio <0.35). Distance-based methods gave results more congruent with morphological sorting than tree-based methods, especially within the Chalcidoidea superfamily.

Taking into account the comparison between the cost and time of molecular and morphological classification methods, our results suggest that mini-DNA barcoding to estimate a proxy for Hymenoptera species diversity is more economical than MorphOTU sorting. However, although MorphOTU and MOTU methods gave different numbers of species, actual diversity analyses take into account abundance and other parameters. We did not evaluate whether MorphOTUs and MOTUs would yield sufficiently equivalent results in actual α- and β-diversity research: that is, they may yet both be fit for purpose.

Keywords: High throughput DNA barcoding, Hymenoptera, biodiversity, species delimitation, COI.

Outline

List of tables

Chapter 2

Chapter 3

List of figures

Chapter 1

Chapter 2

Chapter 3

List of Abbreviations

ABGD: Automatic Barcode Gap Discovery

AR: absorbance ratio

ASAP: Assemble Species by Automatic Partitioning

BINs: Barcode Index Numbers

BOLD: Barcode of Life Database

bp: Base Pairs

bPTP: Bayesian Poisson Tree Processes

BSA: Bovine Serum Albumin

COI: Cytochrome C Oxidase Subunit I

CTAB: Cetyltrimethyl Ammonium Bromide

ddPCR: Droplet Digital PCR

dPCR: Direct polymerase Chain Reactions

F: Forward

GMYC: Generalised Mixed Yule Coalescent

h: Hour

HTS: High-Throughput Sequencing

Kb: Kilobase

Mb: Mega base

MCMC: Monte Carlo Markov Chains

min: Minutes

ML: Maximum Likelihood

mL: *Milliliter*

mm: Millimeters

MorphOTU: Morphological Operational Taxonomic Unit

MOUT: Molecular Operational Taxonomic Units

ng: Nanogramme

NJ: Neighbor-Joining

(P): Prior maximum divergence of intraspecific diversity

p: p-value

PacBio: Pacific Biosciences

PCR: Polymerase Chain Reaction

PE: Paired End

R: Reverse

s: Seconde

SDS: Sodium Dodecyl Sulfate

SMRT: Single-Molecule Real-Time

Tb: Terabyte

V: Volte

W: Shapiro-Wilk

X: gap width

ZMWs : Zero-Mode Waveguides

µL: Microlitre

µm: Micrometre

µM: Micromolaire

Dedications

To my supportive, loving parents along with the rest of my family who supported me throughout these years and made it possible for me to achieve this milestone in my academic career, without you I wouldn't imagine myself writing this today.

To my friends with whom I witnessed good company and support.
To my delicate self who gave me the patience, patience and patience to accomplish this work.

Acknowledgements

I'm grateful to God who created every detail in my life that lead me to this point.
I want to acknowledge everyone who gave a hand, contributed and helped, even with a word, in accomplishing this work, I'll never forget your help and kindness, thank you very much, from the bottom of my heart.

I want to specifically thank my **Supervisor**, Professor **Colin Favret,** professor at the University of Montreal and the head of entomological laboratory at IRBV, who offered me the chance to work in this field with great support and kindness, and who was always available to guide my steps throughout the entire internship. I want to thank Professor Favret for giving me the opportunity to be a part of its laboratory's members.

Member of the jury, **Professor Étienne Léveillé-Bourret** and **Professor Simon Joly,** professors at the University of Montreal, I'm honored and thankful that you accepted the call for judging my work. I thank you for your time and effort for bringing your knowledge and experience to evaluate my work.

I want to particularly thank **Dr. Romain Nattier,** Associate Professor at The Institute of Systematics, Evolution, Biodiversity, France, who guided, advised and supported me with the part of bioinformatic analyses

Chapter1

Introduction and literature review

General introduction

Rapid, cost-effective and reliable identification of a large number of specimens is paramount in many aspects of biological research, especially for studies in systematics, ecology, evolutionary biology, conservation biology, biodiversity and biomonitoring (Misa *et al*., 2009; Yu *et al*., 2012; Hirai *et al*., 2015; Miller *et al*., 2016; Piper *et al*., 2019; Morinière *et al*., 2016; Janzen *et al*., 2020). Correct species identification not only allows access to the literature on a taxon or certain estimation of species richness (Rivera and Currie, 2009; Hebert *et al*., 2016), but also permits the implementation of control measures for species of medical or agricultural importance (Reuter *et al*., 2015; Arje *et al*., 2020). Misidentifications can lead to inadequate control measures with negative socioeconomic implications and reduced ability to predict threats under environmental degradation such as climate change (Zhou *et al*., 2014; Arje *et al*., 2020).

Threats of biodiversity loss induced by anthropogenic change, the vast number of undescribed species (Ratnasingham and Hebert, 2013; Hebert *et al*., 2016; Bennett *et al*., 2019a), the prevalence of mis-identification (Derraik *et al*., 2002; 2010, Pearson *et al*., 2011) and the ongoing loss of taxonomic expertise, known as the "taxonomic impediment", globally (Hoagland, 1995; Derraik *et al*., 2002, Wheeler *et al.,* 2004, Pearson *et al*., 2011) and especially in Canada (Council of Canadian Academies and Expert Panel on Biodiversity Science, 2010, Hebert *et al*., 2016), have spurred the scientific community to search for more efficient methods of specimen identification (Godfray, 2002; Monaghan *et al.,* 2005; Cywinska *et al.,*2006; Derraik *et al*., 2010; Hernández *et al*., 2012; Bennett *et al*., 2019a).

However, traditional biodiversity assessments based on morphological identification typically require substantial training (Derraik *et al*., 2002; Jinbo *et al*., 2011) and are time-consuming and expensive. In fact, Carbayo and Marque (2011) estimated that a comprehensive inventory of animal species using morphology would require $250 billion and another 600 years, and may not always provide resolution to the species level (Cywinska *et al.,* 2006; Packer *et al.,* 2009, Derraik *et al*., 2010; Hebert *et al*., 2016; Loit *et al*., 2019).

In order to accelerate and facilitate the process of species identification, many newer methods use molecular tools to classify and identify organisms (Hajibabaei *et al*., 2006, Packer *et al.,* 2009, Borisenko *et al*., 2009; Leite *et al*., 2012; Hebert *et al*., 2016; Bennett *et al*., 2019a). DNA barcoding, a taxonomic method that uses a short DNA sequence (a set of 658 nucleotides in the 5' region of the mitochondrial cytochrome oxidase c subunit 1 gene (COI) that has the right

level of variability to usually be diagnostic at the species level) (Hajibabaei *et al*., 2006, Hebert *et al*., 2003; Ivanova *et al*., 2009; Bahder *et al.*, 2015; Martoni *et al*., 2019; Srivathsan *et al*., 2019), has gained increased attention and acceptance from members of the scientific community interested in documenting the Earth's biodiversity, especially that of animals (Hebert *et al*., 2003; Savolainen *et al*., 2005, Hajibabaei *et al*., 2006, Borisenko *et al*., 2009; Ivanova *et al*., 2009, Janzen *et al*., 2009, Strutzenberger *et al*., 2010, Leite *et al*., 2012; Hebert *et al*., 2016; Dopheide *et al*., 2019).

DNA barcoding has been postulated as a viable tool to enhance taxonomic research by discovering new taxa and verifying morpho-taxonomic hypotheses, responding to pressing biodiversity needs (Smith *et al*., 2005; Hebert *et al*., 2013; Muhammed Taher and Akthar, 2016) as well as a tool for rapid identification (Giantsis *et al*., 2015; Grosdidier *et al*., 2017; Martoni *et al*., 2019). However, due to several problems, including its three main steps, DNA extraction, PCR amplification, sequencing and data analysis, DNA barcoding has been the subject of a vigorous debate in the scientific community between those embracing it (Kress and Erickson, 2008; Schindel and Miller, 2005; Leite *et al*., 2012; Bahder *et al.*, 2015), and those opposing the way its most fervent supporters propose using it (Mallet and Willmott, 2003; Ebach and Holdrege, 2005; Hickerson *et al.* 2006; Vogler and Monaghan, 2007).

Several of the biggest challenges of DNA barcoding are related to DNA extraction methods that can be destructive (Hoff Olsen *et al*., 1999; Ivanova *et al*., 2006; Hunter *et al*., 2008; Musapa *et al*., 2013; Bahder *et al.*, 2015), expensive (Ivanova *et al*., 2006, Guzmán-Larralde *et al*., 2017; Giantsis *et al*., 2015), time consuming (Favret, 2005; Miura *et al*., 2017, Guzmán-Larralde *et al*., 2017; Suaste *et al*., 2019), unreliable (Martoni *et al*., 2019) and toxic (Gilbert *et al*., 2007; Nancy *et al*., 2010). Additionally, PCR-based detection with specific oligonucleotide primers lack the capacity to detect species or strains other than those targeted (Castalanelli *et al*., 2010) or worse, yield false positive signals (Grosdidier *et al*., 2017). Furthermore, the cost of obtaining DNA barcodes via Sanger sequencing is prohibitive when thousands of specimens have to be processed (Missa *et al*., 2009; Yu *et al*., 2012; Shokralla *et al*., 2014, Reuter *et al*., 2015; Lanner *et al*., 2019; Loit *et al*., 2019).

Insects are the most diverse class of the animal kingdom (Chapman, 2009), with more than one million described species (Mora *et al*., 2011; Hofreiter *et al*., 2015; Stork.,2018; Patzold *et al*., 2020) and multiple millions yet to be (Gilbert *et al*., 2007; Zhang, 2011; Stork.,2018; Yeo *et al*., 2020). Entomologists cannot afford to sort all specimens in traps via barcodes through

Sanger sequencing. For instance, based on the pricing listed by the Canadian Centre for DNA Barcoding (https://ccdb.ca/pricing/), Sanger sequencing can exceed 23CAD per specimen. Thus, identification currently demands substantial resources to deal with an incredible diversity of species and a large number of individuals (Favret and Voegtlin, 2001; Sheikh *et al*., 2016; Favret *et al*., 2019; Seibold *et al*., 2019; Wagner, 2020).

Regarding the high cost, most studies on insect biodiversity assessment begin with pre-sorting samples based on morphology (often done by parataxonomists) to define Morphological operational taxonomic units (MorphOTUs) also called morphospecies (distinct taxa distinguished by readily discernible morphological traits (Derraik *et al*., 2001)), then applying DNA-barcoding for few samples to test the effectiveness of morphological identification (Barrett and Hebert, 2005; Renaud *et al*., 2012; Riedel *et al*., 2013; Hebert *et al*., 2016; Knox *et al*., 2020). However, this workflow can be problematic especially with cryptic species that are morphologically indiscernible (Wang *et al*., 2016; Stork *et al*., 2018), which lead to incorrect evaluation of biodiversity.

Due to the necessity for a cheaper and faster High throughput DNA barcoding approach, I aim to overcome the aforementioned disadvantages by targeting the three main steps including DNA extraction, amplification, sequencing, and data analysis, combining the DNA barcoding approach, originally developed to identify single specimens, with high-throughput sequencing (HTS) technologies. Taking into account the existing number of undescribed species, the "taxonomic impediment", the difficulty of identification, and the lack of taxonomists, there is a need to compare the effectiveness of molecular vs morphological data for the purpose of biodiversity assessment. For our research, we take the Hymenoptera insect group as a model.

1. Hymenoptera biodiversity

Hymenoptera is one of the largest insect orders, with approximately 160,000 species in more than 8420 genera from 123 extant families in the word fauna (Königsmann, 2008, Aguiar *et al*., 2013; Forbes *et al*., 2018). A Canadian fauna of 8757 species from 83 families has been recorded (Bennett *et al*., 2019a). Between 1979 and 2018, the recorded Canadian fauna increased by approximately 46%, or 5322 species (Masner *et al*., 1979; Bennett *et al*., 2019a). Ontario is first in Canada, accounting for 57.5% of all species, followed by Quebec (4207, 45.5%) and British Columbia (4063, 43.9%). The Ichneumonoidea is the largest super-family of the order Hymenoptera, both in the world (>48,000 valid species), accounting for 33% of all Hymenoptera (145,000 species) and 3% of all known life (Peters *et al*., 2017; van Achterberg *et al*., 2017; Sharanowski *et al*., 2021), and in Canada (4202 species) (Yu *et al*., 2016). Ichneumonidae is the most speciose family in Canada with 3037 species (Aguiar *et al*., 2013; Bennett *et al*., 2019), followed by Braconidae, the sister group of Ichneumonidae (Heraty *et al*., 2011; Sharkey *et al*., 2012; Peters *et al*., 2017; Branstetter *et al*., 2017; Bennett *et al*., 2019b).

In addition to the Ichneumonoidea, Canada's Hymenoptera faunal structure is dominated by 16 families of Chalcidoidea, with more than 22,700 described species (Huber, 2017), and Apoidea 1352 species: 15.4%). These three super-families account for over three-quarters of the recorded named species (77.3%) (Bennett *et al*., 2019a).

Despite their fundamental role as parasitoids, predators, pollinators and bio-indicators in all terrestrial ecosystems (Sheffield *et al*., 2009; Peters *et al*., 2017), many families of Hymenoptera in the world (Forbes *et al*., 2018; Belokobyl'skij and Lelej, 2019) and in Canada are inadequately known (Peters *et al*., 2017). Over 10,000 undescribed species were estimated to exist in Canada using Barcode Index Numbers (BINs) (Ratnasingham and Hebert, 2013) based on a 2% or greater DNA barcode sequence divergence (Bennett *et al*., 2019a).

Their small size and a paucity of taxonomic resources make the parasitoids poorly known (Rasnitsyn *et al*., 1988; LaSalle, 1993; Bennett *et al*., 2019b). In fact, through a simple biological model studying the global ratio of wasp parasitoids to hosts (P:H), Forbes e*t al*. (2018) have shown that the Hymenoptera are almost certainly the most speciose animal order, with parasitoid wasps alone constituting 2.5–3.2 time more species than the Coleoptera (beetles): the ratio of genus-specialist parasitoids to hosts was estimated to be close to one but,

additively most coleopteran species are attacked by more than one hymenopteran parasitoid species (Forbes *et al*., 2018).

Current studies of Favret *et al*. (2019), working with Voegtlin-style suction traps, have shown the exceptional diversity of Hymenoptera that are under-documented in Canada (Hebert *et al*., 2016; Bennett *et al*., 2019a). Taking into account the importance of Hymenoptera in the environment and their role as specialist parasitoids of other insects, they may serve as representative for insect diversity more generally (New, 2011; Favret *et al*., 2019; Morinière *et al*., 2019).

As mentioned previously, Hymenoptera include the smallest insects, like mymarommatoids, a fraction of a millimeter in length (Bennett *et al*., 2019b). Other super-families, such as Diaprioidea and Platygastroidea are heavily sclerotized (Belokobyl'skij and Lelej, 2019), giving low DNA yields in terms of both quality and quantity (Wang *et al*., 2019; Ulmer *et al*., 2021). Additionally, the DNA barcodes of some Hymenoptera groups are notoriously difficult to acquire (Kaartinen *et al*., 2010; Cruaud *et al.*, 2019; Vasilita *et al*., 2022). Appropriate DNA extraction, barcode amplification, and sequencing protocols must be usable with the full range of Hymenoptera size, sclerotization and genomic variability.

What follows is a literature review of the modern methods of DNA extraction, PCR amplification and high-throughput sequencing (HTS) platforms that could expand the current capability of DNA barcoding approaches.

2. DNA meta-barcoding workflow

We here refer to meta-barcoding as the use of high-throughput sequencing (HTS) for generating multiplexed barcodes of separate specimens; such an approach is needed for lowering the cost of barcodes when thousands of specimens are processed (Yu *et al*., 2012; Ji *et al*., 2013; Liu *et al*., 2013; Zhang *et al*., 2018; Srivathsan *et al*., 2019) (Figure 1 addenda). This process is also called "sample multiplexing": individual samples are pooled and sequenced simultaneously during a single run on a HTS platform. Meta-barcoding or sample multiplexing has the advantage of processing many samples without significantly increasing cost and time. Molecular tags are added to each DNA fragment during HTS library preparation, so that each read may be identified and tracked back to a single original template (Ståhlberg *et al*., 2017; Illumina, 2022).

Routine DNA meta-barcoding studies inevitably start with either field sampling or assessment of the suitability of existing samples (Patrick *et al*., 2016; Deagle *et al*., 2019). Careful

consideration of sampling and curation procedures is needed to avoid DNA contamination and ensure DNA preservation (Jeunen *et al*., 2018; Erdozain *et al*., 2019). The next process consists of multiple laboratory steps: DNA extraction, PCR amplification, and DNA sequencing. For reliable, cost effective and fast assessment of insect biodiversity, sufficient technical knowledge and informed choice are required at each step.

2.1. DNA extraction

DNA extraction is a routine step in the DNA barcoding process. A variety of methods have been established to isolate DNA molecules from insects (Favret, 2005; Ivanova *et al*., 2006; Chen *et al*., 2008; Musapa *et al*., 2013; Asghar *et al*., 2015; Guzmán-Larralde *et al*., 2017; Suaste *et al*., 2019) and many DNA extractions kits are commercially available (Hernandez *et al*., 2012; Bahder *et al*., 2015; Giantsis *et al*., 2015; Miura *et al*., 2017; Patzold *et al*., 2020). The most common insect DNA extraction procedures necessitate maceration of the material, destroying the morphological characteristics required for identification (Ivanova *et al*., 2006; Chen *et al*., 2010; Musapa *et al*., 2013), resulting in the loss of the physical reference specimen and representing a major drawback for museum specimens (Muhammed Taher and Akthar, 2016; Rohland *et al*., 2018; Patzold *et al*., 2020). Preservation of insect reference specimens is critical to ensure traceability between the molecular and morphological features, especially in the case of taxonomic reassignments (Castalanelli *et al*., 2010; David *et al*., 2013; Miura *et al*., 2017; Martoni *et al*., 2019; Piper *et al*., 2019). For this laboratory process, my goal is to use a non-destructive reliable, fast and economic DNA extraction method to deal with a large number of samples.

There are five basic steps of DNA extraction: lysis, protein removal, washing, and elution (Hoff-Olsen *et al*., 1999; Ivanova *et al*., 2006; Giantis *et al*., 2015). These are consistent across DNA extraction protocols, yet DNA of interest can be isolated using a variety of different chemistries (Phenol-Chloroform DNA extraction, salting out extraction, spin Column DNA separation, chelation extraction, magnetic beads extraction and heat extraction). Each has a characteristic binding capacity and each has its unique advantages and disadvantages (Table 1 addenda). My choice of the DNA extraction technique is based on several criteria notably cost, speed, yield, capacity to be amplified and degree of toxicity. I propose that these criteria be a tool to evaluate the effectiveness and adaptability of each method for my objectives. When the efficiency is adequate, cost and time are my primary benchmarks.

2.1.1. Organic phenol-chloroform DNA extraction

In the last decade, a number of phenol–chloroform DNA extraction methods have been published. Based on a liquid-liquid extraction, these allow at least a partial preservation of the morphological features of the specimen (Favret, 2005; Gilbert *et al*. 2007; Rowley *et al*. 2007; Castalanelli *et al*., 2010; Porco *et al*. 2010; Bahder *et al*., 2015). The phenol: chloroform mixture is immiscible with water. Centrifugation will cause two distinct phases to form in an extraction tube: an upper aqueous phase containing isolated DNA, and a lower organic phase that contains lipids and cellular debris (Sun, 2010; McKiernan and Danielson, 2017). Many techniques based on phenol-chloroform extraction have been using to extract DNA from insects, mostly differing in the sodium dodecylsulfate (SDS) lysis step that can be based on (Chen *et al*., 2010; Wang *et al*., 2019; Murthy *et al*., 2022), CTAB (Chen *et al*., 2010; Nancy *et al*., 2010; Murthy *et al*., 2022) or CaCl2 buffer (Gilbert *et al*., 2007; Suaste *et al*., 2019). These methods were efficient to extract DNA from a wide range of insects such as xylophagous species (Nancy *et al*., 2010), small samples (Wang *et al*., 2019), heavily sclerotized insects like Coleoptera and Neuroptera (Asghar *et al*., 2015; Guzmán-Larralde *et al*., 2017). However, phenol-chloroform methods are relatively time-consuming, cause health risks and possible high rates of DNA loss (Chen *et al*., 2010; McKiernan and Danielson, 2017) (Table 1 addenda).

2.1.2. Salting out extraction

As an alternative to the phenol-chloroform extraction procedure, the salting-out method has the advantage of not using toxic chemicals (Chen *et al*., 2010; Sun, 2010; Evans *et al*., 2013). It is based on the principle that proteins and other cellular contaminants precipitate in a saturated salt solution, due to their relative hydrophobicity, while DNA does not (Sun, 2010). One kit based on this organic DNA extraction method is The Master Pure™ Complete DNA and RNA Purification Kit (Biosearch technologies, London, UK) which permits rapid purification of high-molecular-weight nucleic acids from many samples that are processed simultaneously in 30 minutes (The Master Pure™ Complete DNA and RNA Purification Kit, 2012; Evans *et al*., 2013). However, as with phenol-chloroform chemistry, the salting out extraction is a two-step process involving transfer of reagents between tubes, increasing the risk of contamination.

2.1.3. Spin column DNA separation

A number of commercial kits using a centrifuged separation column have been developed for rapid and efficient isolation of genomic DNA (Théry *et al*., 2017; Selleres *et al*., 2018; Oppert *et al*., 2019; Patzold *et al*., 2020). It is a solid phase extraction method relying on the fact that

nucleic acid will bind, under certain conditions to a silica gel membrane inside the spin column (Hoyt *et al*., 2001).

The most commonly used DNA extraction kit is Qiagen DNeasy Blood and Tissue (Qiagen, Hilden, Allemagne) (Johnson *et al*., 2001; Chen *et al*., 2010; Giantsis *et al*., 2015; Djurhuus *et al*., 2017; Miura *et al*., 2017; Guzmán-Larralde *et al*., 2017; Théry *et al*., 2018). It is designed for rapid purification of total DNA from a variety of sample sources including fresh or frozen tissues and cells, blood or bacteria (DNeasy Blood & Tissue Handbook, 2020). Many researchers have adopted this method to extract DNA from a large range of insects. Some studies employed DNeasy kits for non-destructive DNA extraction from small-sized terrestrial arthropods, dry-preserved insects commonly held in entomology collections, and other soft-bodied small arthropods (e.g., Djurhuus *et al*., 2017; Guzmán-Larralde *et al*., 2017; Miura *et al*., 2017; Théry *et al*., 2018, Suaste *et al*., 2019; Wang *et al*., 2019; Liu *et al*., 2020).

This kind of chemistry involves all DNA extraction steps with numerous variations in lysis incubation temperature (37°C-56°) and time (1–36 hours) (Chen *et al*., 2008; Guzmán-Larralde *et al*., 2017; Miura *et al*., 2017 Santos *et al*., 2018; Suaste *et al*., 2019). DNA is purified using the DNeasy spin-column in as little as 20min. Precipitated DNA is selectively bound to the column membrane in the presence of high concentrations of chaotropic salt, as dissolved contaminants pass through. In two more wash processes, any remaining contaminants and enzyme inhibitors are removed, and DNA is eluted in water or buffer, ready to use (DNeasy Blood & Tissue Handbook, 2020). The DNeasy kit can extract DNA safely and easily (Giantsis *et al*., 2015; Hartop *et al*., 2020), preserving morphological features of the specimens (Giantsis *et al*., 2015; Guzmán-Larralde *et al*., 2017; Suaste *et al*., 2019; Hartop *et al*., 2020), but it can become relatively expensive when many samples need to be processed, and long incubations (sometimes more than 20 hours) incur a high investment in time. Other kits using silica column-based approach were investigated (Oppert *et al*., 2019; Patzold *et al*., 2020), such as the gSYNC (Geneaid, New Taipei, Taiwan), E.Z.N.A. Insect DNA Kit (Omega BioTek, Norcross, USA) and PCR & DNA Clean-up Kit (Biolabs, New England, USA).

The high success rate of spin column DNA separation methods (Liu *et al*., 2020) allows them to be a control method during our research, but the time and cost factors favor methods that allow to process a large number of samples.

2.1.4. Heat extraction

Other methods that do not require cumbersome columns or phenol and chloroform, or other caustic solvents, based on heat treatment, are very efficient and promising. To discover phorid diversity, Srivathsan *et al*. (2019) applied a workflow process using a simplified DNA extraction protocol on 8700 specimens that requires only heat treatment through incubation steps to lyse the tissue material, release the DNA, and degrade compounds inhibitory to amplification. This procedure is fast, simple, inexpensive (reagent costs 0.06 per specimen) and is easily scaled to process hundreds of samples in multi-well plates using robotic automation. Thus Wang *et al*. (2019) recommended to use this relatively cheap and rapid DNA extraction can be done using QuickExtract™ DNA extraction solution Tissue (Lucigen, New York, USA) (Quick Extract™ DNA Extraction Solution, 2018) to fit any form of biodiversity study.

Another method based on heat treatment with an alkaline lysis buffer is the Hot Sodium Hydroxide and HotSHOT (UConn Health, Farmington, USA) method. It requires the addition of Alkaline Lysis Reagent (NaOH and disodium EDTA), an incubation step at 95°C followed by neutralization with Tris-HCl. Its main advantage is that it can easily be completed in less than an hour (20-40 min) (Truett *et al*., 2000; Suaste *et al*., 2019, Patzold *et al*., 2020). Furthermore, the PrepGEM kit (Promega Corporation, Madison, USA) is an easy extraction method, requires a simple digestion step followed by incubation at 75˚-95°C for 5-15min. It is cheap and fast for freshly preserved specimens, the total time required is 20min, and high quality DNA can be extracted from many types of material using only a single tube (Asghar *et al*., 2015).

2.1.5. Chelation extraction

Chelex resin extraction (BioRad, Californie, USA) is based on the addition of chelating ion exchange resins that act to bind polyvalent metal ions such as magnesium. Heat is applied to lyse the cells and releasing DNA, while the chelating ion exchange resin protects the DNA from degradation (Walsh *et al*., 1991). As a non-destructive DNA extraction method, the Chelex method is useful (Cornils, 2015; Musapa *et al*., 2013; Murthy *et al*., 2022) and merits being used more frequently, especially in the case of tiny insects for which the whole body is needed in order to realize an appropriate yield (Musapa *et al*., 2013; Miura *et al*., 2017). This method can be easily automated and requires minimal sample transfer, decreasing the opportunity for contamination or other sample mishandling (McKiernan and Danielson, 2017). Compared with spin-column-based methods Chelex-based DNA extraction methods are cheaper and faster

(37min), but yields are inconsistent in quality and quantity (Casquet *et al*., 2011; Cornils, 2015) and may fail to remove PCR inhibitors (Lagisz *et al*., 2010).

2.1.6. Magnetic beads extraction

The ChargeSwitches technique (Invitrogen, Waltham, USA) is a simple method consisting in the addition of magnetic beads in solution that, at low pH, attract DNA (Asghar *et al*., 2015). Once contaminants have been removed, increasing the pH to 8.5 neutralizes the ChargeSwitch magnetic beads and the DNA is eluted (Sun, 2010; Asghar *et al*., 2015). This method has a slightly lower yield than the other solid-phase adsorption methods (e.g., Chelex) (Sun, 2010), possibly because buffers are not entirely eliminated in several steps in order avoid disrupting the magnetic particles. Nevertheless, with no need for centrifugation or a vacuum manifold, the procedure is simple (Asghar *et al*., 2015).

In conclusion, the appearance of new high-throughput sequencing technologies also requires the development of high-throughput methods of DNA extraction (Liu *et al*., 2020; Patzold *et al*., 2020). Such methods have utilized automatic liquid-handling robots (Ivanova *et al*., 2006), sonicators (Hunter *et al*., 2008), and 96-well plates (Lagisz *et al*., 2010; Dentinger *et al*., 2010). The major problem when using 96-well plates with a silica column-based approach is the large size of the columns. The transition of these protocols to a phenol-chloroform approach or other simple procedure previously mentioned would help to overcome this restriction, although ordinarily the use of toxic compounds should be avoided and phenol may be harmful to some components of a liquid-handling robot. There is also an opportunity for developing a fast, simple and inexpensive extraction method that can yield high-quality DNA from insect tissues by combining a traditional cheap single-tube method like Chelex extraction and more recent high-throughput methods in order to overcome the several drawbacks of the kits.

2.2. Direct PCR

DNA extraction may be avoided altogether through the use of direct PCR. In fact, the time and cost for obtaining DNA barcodes is reduced by placing tissue directly into a PCR master mix without DNA extraction prior to amplification of the target gene (Rochlin *et al*., 2007; Wong *et al*., 2014; Meier *et al*., 2016; Guzmán-Larralde *et al*., 2017; Wang *et al*., 2018). Direct PCR reduces the overall time taken (by 4-16 hours, depending on extraction protocol) and the cost by eliminating the usage of kits and decreasing manpower needs (Herandez *et al*., 2012; Meier *et al*., 2016). Direct PCR has been known for more than 20 years (Panaccio *et al*., 1993), but

success rates have either been unreported or low (Panaccio *et al*., 1993; Grevelding *et al*.,1996; Rochlin *et al*., (2007; Wong *et al*. 2014; Meier *et al*., 2016).

Alternatively, direct PCR can be used as long as users adapt the pipeline to their own insect group (Wong *et al*., 2014), optimizing reagent amounts, template tissue volume, and cycling conditions for boosting direct PCR success rates. Additionally, obtaining amplicons with direct PCR is cost-effective (0.16 US$) (Yeo *et al*., 2018) and sufficiently simple that even inexperienced personnel can execute it (Wang *et al*., 2019). Direct PCR in combination with meta-barcoding can help overcome taxonomic bias against small specimens that parataxonomists tend to disregard (Stribling *et al*., 2008; Orlofske and Baird, 2013). Yet the tissue sampling of specimens for direct PCR does cause specimen damage. Occasionally, an additional tissue sample is required. As a consequence, some researches argue that direct PCR should be used only for taxa with large numbers of very similar specimens (Meier *et al*., 2016).

2.3. PCR Amplification

2.3.1. Marker enrichment

Amplification of DNA first requires the appropriate selection of a taxonomic marker or barcode locus. This is a critical first step in design of DNA barcoding assay because all downstream species detection and identification will rely on how conserved this marker is across taxa, and the discriminatory power of the nucleotide variation contained within it (Yu *et al*., 2012; Freeland *et al*., 2017; Pipper *et al*., 2019). The marker widely adopted for animal DNA barcoding is the mitochondrial gene for cytochrome oxidase I (COI) (Meusnier *et al*., 2008; Leite *et al*., 2012; Andújar *et al*., 2018; Elbrecht *et al*., 2019). In fact, the choice of a mitochondrial gene as a universal barcode was mostly driven by the fact that the mitochondrion is maternally inherited, avoiding problems with recombination (Hlaing *et al*., 2009; Leite *et al*., 2012). Also, mitochondrial genes have a high copy number and a high mutation rate when compared with most nuclear markers, which results in high degrees of inter-specific polymorphism and divergence (Leite *et al*., 2012; Choi *et al*., 2018).

Metabarcoding studies on bulk collections of animals usually target a subset of the cytochrome c oxidase subunit I (COI), particularly 658bp at the 5’ end of the “Folmer” region (Folmer *et al*., 1994; Yu *et al*., 2012; Andújar *et al*., 2018; Elbrecht *et al*., 2019; Srivathsan *et al*., 2019). This gene region has gained broad adoption because of a rapidly expanding reference database, particularly well-represented in the Barcode of Life Database (BOLD-Ratnasingham and Hebert, 2007; Porter and Hajibabaei, 2018) and its good taxonomic resolution (Meusnier *et al*., 2008).

Other genes like the gene Gnd of the aphid obligate bacterial endosymbiont *Buchnera aphidicola*, the mitochondrial gene ATP6 (Chen *et al*., 2013, Lee *et al*., 2014) and many nuclear genes (Simon *et al*., 2010) were successfully tested. They are often useful in phylogenetic analyses of higher-level arthropod taxa (Caterino *et al*., 2000, Simon *et al*., 2010; Depa *et al*., 2017). For example, the nuclear gene EF-1a has been employed in the phylogeny reconstructions in several groups of insects (Cho *et al*., 1995, Kim and Lee, 2008; Condamine *et al*., 2013, Lin *et al*., 2013, Cooper *et al*., 2014; Théry *et al*., 2017).

Despite the effectiveness of other possible barcode markers, many taxa currently only have COI sequence data publicly available due to the aforementioned particularities of COI (Liu *et al*., 2020; Piper *et al*., 2019).

2.3.1.1. Mini-barcodes

Due to the limitations in the size of DNA fragments sequenced by HTS platforms (Binladen *et al*., 2007; Shokralla *et al*., 2014; Brandon-Mong *et al*., 2015), metabarcoding has typically been restricted to targeting short fragments of the COI DNA barcode, named mini-barcodes: these have several advantages. Firstly, such amplicons are easier to obtain when the DNA in the sample is degraded: the idea of mini-barcodes had been investigated in the context of degraded DNA samples (Smith *et al*., 2005; Hajibabaei *et al*., 2006), suggesting that singular COI barcodes of sizes between 135 and 250 bp can reliably distinguish most animal species (Hajibabaei *et al*., 2006; Meusnier *et al*., 2008; Hajibabaei *et al*., 2011; Leray *et al*., 2013). Secondly, mini-barcodes can be sequenced at low cost using tagged amplicon sequencing on short-read sequencing platforms (e.g., Illumina) (Bentley *et al*., 2008; Guo *et al*., 2008; Reuter *et al*., 2015; Lanner *et al*., 2019). Finally, mini-barcode primers are accessible for a wide range of arthropod groups with a substantial number of species (Meusnier *et al*., 2008; Hebert *et al*., 2013; Little, 2014).

Published tests of mini-barcodes comparing their performance with full-length barcodes yielded conflicting results, but most of them demonstrated no significant difference in performance for species identification between full-length and mini-barcodes as long as they are of moderate length (>200-400bp) (Meusnier *et al*., 2008; Wang *et al*., 2018; Yeo *et al*., 2020). Only very short mini-barcodes (<200bp) perform poorly, especially when they are located near the 5’ end of the Folmer region.

2.3.2. Primers

The precise primer set to use will be determined by the study's context, amplicon length requirements, and desired taxonomic resolution (Meusnier *et al*., 2008; Porter and Hajibabaei, 2018). In general, primers should preferentially target hypervariable DNA regions: these permit high resolution taxonomic discrimination for which extensive libraries of reference sequences are available (Huber *et al*., 2009; Dopheide *et al*., 2019).

The majority of DNA-barcoding and metabarcoding studies on insects employed the universal primers LCO1490 and HCO2198 (Folmer *et al*., 1994; Stahlhut *et al*., 2013; Dopheide *et al*., 2019; Srivathsan *et al*., 2019) for the full-length DNA barcode (658bp). In fact, Srivathsan *et al*. (2019) recovered high PCR success rates in Diptera using these primer pairs. Additionally, Stahlhut *et al.* (2013), used LCO1490 and HCO2198 to identify molecular operational taxonomic units (MOTUs) for 7870 Hymenoptera specimens collected in a sub-Arctic environment. Other authors improved on that by using more specific primer pairs, as recommended to increase success rates (Hebert *et al*., 2003; Marien *et al*., 2018).

To assess the German Diptera fauna, PCR-amplified with a cocktail of standard and modified 'Folmer 'primers CLepFolF and CLepFolR for the barcode fragment, revealed 1,735 dipteran MOTUs with a sequence identity higher than 97% to a dipteran record (Morinière *et al*., 2019). Primers with fewer template–primer mismatches were found to be better for quantitative DNA metabarcoding by Piol *et al*. (2018), especially for species with higher relative abundance in a sample.

Furthermore, primers should preferentially target short DNA fragments (e.g., < 400bp) to maximize richness estimates and increase the probability of recovering DNA templates that are more degraded such as samples preserved for extended periods of time (Deagle *et al*., 2006; Leray *et al*., 2013). Many researchers used m1COlintF (Leray *et al*., 2013) and modified jgHCO2198 (Geller *et al*., 2013; Meier *et al*., 2016) to amplify 313bp to evaluate the Diptera and Hymenoptera fauna (Leray *et al*., 2013; Wang *et al*., 2018; Yeo *et al*., 2018; Morinière *et al*., 2019; Srivathsan *et al*., 2019). These primers performed well across arthropod diversity, with higher success rates than versatile primer sets traditionally used for DNA barcoding (Leray *et al*., 2013; Morinière *et al*., 2019; Elbrecht *et al*., 2019).

Finally, it is helpful to evaluate the effectiveness of primer sets at the beginning of a project by *in vitro* tests with mock communities (Elbrecht and Leese, 2015; Brandon-Mong *et al*., 2015; Leray and Knowlton, 2015) or by *in silico* tests (Clarke *et al*., 2014; Elbrecht and Leese, 2016; Piñol *et al*., 2018). These tests ensure that primer sequences are appropriate for the underlying target community.

2.3.3. PCR conditions

While primer choice is critical for metabarcoding projects, PCR can also be biased by the polymerase used (Nichols *et al*., 2018), the number of thermocycles (Vierna *et al*., 2017; Krehenwinkel *et al*., 2016), template GC content (Braukmann *et al*., 2019), inhibitors (Sellers *et al*., 2018), and annealing temperature (Clarke *et al*., 2017; Krehenwinkel *et al*., 2018; Elbrecht *et al*., 2019). It is generally assumed that primers bind better at lower annealing temperatures leading to better taxonomic recovery (Aylagas *et al*., 2016; Elbrecht *et al*., 2019). Nonetheless, at very low anneleaing temperature, primers may bind nonspecifically to the template (Ishui and Fukui, 2001). Optimization steps are required.

Furthermore, templates with suboptimal GC contents can be disfavored during amplification. In fact, results of Nichols *et al*. (2018) indicated that GC bias can confound metabarcoding-based study, although some polymerases are known to perform well with sequences of specific GC content (Braukmann *et al*., 2019).

Many authors have demonstrated that the type of polymerase influences PCR success rates: the highest success rates were achieved using TaKaRa ExTaq (Thermo Fisher Scientific, Waltham, MA)), but high success rates can be attained with low-cost or even homemade enzymes (Meier *et al*., 2016). However, HotStar Taq (Qiagen, Hilden, Allemagne) did not yield successful amplifications (Wong *et al*., 2014). Given that simultaneous optimization of tissue quantity and enzyme is time-consuming, researchers recommend first optimizing the former while using a high-fidelity Taq polymerase. Once PCR success rates are high, a cheaper and more versatile Taq polymerase can be tested (Wong *et al*., 2014, Meier *et al*., 2015; Wang *et al*., 2019). Finally, the number of PCR cycles has also been shown to influence results: a higher number of PCR cycles might increase the likelihood that rare molecules are amplified (Weyrich *et al*., 2017; Vierna *et al*., 2017).

2.4. High-throughput sequencing platforms

PCR amplification followed by dideoxy chain-termination sequencing, also known as Sanger sequencing (Missa *et al*., 2009; Bik *et al*.; 2012), has been used for the production of nearly all

of the existing content of public DNA barcode libraries (Smith *et al*., 2005; Ivanova *et al*., 2006; Vogler and Monaghan, 2007; Leite *et al*., 2012). However, cost limitations of Sanger sequencing per specimen restrict its ability to be scaled up to deal with millions of specimens (Dimitrov *et al*., 2017; Krehenwinkel *et al*., 2017; 2018a; 2018b; Lanner *et al*., 2019). In addition, Sanger sequencing libraries require multiple steps that can take more than a week and may fail in cases of mixed or otherwise contaminated samples (Hyde, 2013; Aylagas *et al*., 2016; Mardis, 2017).

In order to overcome the limitations of current identification methods for processing large number of specimens, recent studies have looked to high-throughput sequencing (HTS) technologies to allow DNA barcode-based identification to be conducted in a massively parallel manner (Uroz *et al*., 2016; Knief, 2014; Braukmann *et al*., 2019). Second and third generation high-throughput sequencing methods introduce two short sequence (e.g., 6-9 base pairs) tags, one at each end of the fragment, so that sequencing reads can be sorted into the original samples after HTS (Schnell *et al*., 2015; Meier *et al*., 2016; Wang *et al*., 2018; Srivathsan *et al*., 2019).

HTS platforms fundamentally differ in their ways of recording nucleotides. Each with their unique advantages and disadvantages (Tedersoo *et al*., 2018; Winand *et al*., 2019). Furthermore, these methods exhibit substantial differences in throughput, read length, cost, accuracy and technical biases (Knief, 2014; Reuter *et al*., 2015; Tedersoo *et al*., 2018) (Table 2 addenda).

2.4.1. Second-generation HTS technologies

Second-generation HTS technologies are also known as "short-read" sequencing technologies because achievable reads are relatively short (Siqueira *et al*., 2012; Leggett et Clark, 2017). However, vast quantities of overlapping sequencing reads can be generated through their massively parallel set-up (Rennstam *et al*., 2018).

The first commercially available HTS method was 454 pyrosequencing (Roche Diagnostics, Basel, Switzerland) that was developed in early 2000s (Margulies *et al*., 2005). The read length of pyrosequencing is much shorter than the Sanger method but more cost-effective ($1–2.25 per reaction) and less time-consuming (Hsieh *et al*., 2020). The library preparation required emulsion-PCR to clonally amplify adaptor-ligated DNA fragments on the surface of beads (Ruter *et al*., 2015). The pyrosequencing is based on the "sequencing by synthesis" principle, in which the sequencing is performed by detecting the nucleotide incorporated by a DNA polymerase enzyme (Taberlet *et al*., 2012). It relies on light detection based on a chain reaction when pyrophosphate is released (Shokralla *et al*., 2014). This method can generate up to one

million DNA sequences of up to 700 bases each in a single sequencing run (Reuter *et al*., 20 15). Shokralla *et al*. (2014) used 454 pyrosequencing for generating barcodes one specimen at a time for a small number of 190 Lepidoptera. The barcodes are associated with their specimen of origin by a 10-mer multiple identifier tag (MID). However, the cost per DNA barcode remains relatively high, limiting its utilization in metabarcoding programs.

The Illumina (Sollexa, San Diego, California) and Ion Torrent (Life technologies, US, California) sequencing technologies replaced pyrosequencing 454 in the early 2010s (Mardis, 2017), thanks to their greater throughput at lower costs (see Kemler *et al*., 2013; Reuter *et al*., 2015; Braukmann *et al*., 2019). The template preparation and sequencing steps of Ion torrent technology are conceptually similar to the Roche/454 pyrosequencing platform (Rothberg *et al*., 2011). However, unlike pyrosequencing that couples base incorporation with luciferase-based light production, Ion Torrent's semiconductor sequencing measures pH changes induced by the release of hydrogen ions during DNA extension (Rothberg *et al*., 2011). This method is fast (2–8 hr) and uses the cheapest equipment (Taberlet *et al*., 2012; Reuter *et al*., 2015; Piper *et al*., 2019). Nonetheless, insertions and deletions are common errors, and the successful use of Ion Torrent technology can be hampered by short read length (up to 450 bp) and fluctuating sequence quality, limiting its use especially in analysis of soil and plant samples (see Kemler *et al*., 2013). According to some studies, Ion Torrent application in metabarcoding is limited because of its relatively complex workflow and cost, which are only three to four times less than those for Sanger sequencing (Diekstra *et al*., 2015; Craud *et al*., 2017).

To date, the majority of meta-barcoding and multiplex studies, when individual samples were added and pooled for a single run (Ståhlberg *et al*., 2017), have been conducted using the Illumina platform due to its high-quality reads and relatively inexpensive purchase cost (You *et al*., 2012; Knief, 2014; Wong *et al*., 2014; Krehenwinkel *et al*., 2018; Wang *et al*., 2018). Additionally, Illumina technology's multiplexing capabilities are highly developed, allowing for simultaneous sequencing of many samples in a single run, resulting in more cost-effective HTS (Arulandhu *et al*., 2017).

Simple PCR amplification of the target locus followed by an indexing PCR completes the library preparation (Schirmer *et al*., 2016; Mardis, 2017). Bases are read using a cyclic reversible termination strategy that sequences the template strand one nucleotide at a time through progressive rounds of base incorporation (Bentley *et al*., 2008; Dohm *et al*., 2008; Reuter *et al*., 2015).

Illumina currently produces a suite of sequencers (MiSeq, NextSeq 500, HiSeq and NovaSeq series) optimized for a variety of throughputs and turnaround times (Ruter *et al*., 2015; Piper *et al*., 2019). The MiSeq is the most established platform for insect metabarcoding projects (Piper *et al*., 2019), mainly because it provides reasonable sequencing depth, low sequencing error rates, an affordable cost, and run times as low as 4 h (Brandon-Mong *et al*., 2015; Reuter *et al*., 2015; Liu *et al*., 2020). It produces up to 25 million paired-end reads with lengths of ~300 bp (Schirmer *et al*., 2016). MiSeq can yield 10 000–15 000 barcodes at an NGS cost of 0.14–0.21 USD per barcode (Meier *et al*., 2016).

Some researchers recommend the 'Hiseq' platform because the cost per sample may be impractical for the 'MiSeq' platform for a large number of specimens (Tedersoo *et al*., 2017; 2018; Meier *et al*., 2016; Singer *et al*., 2019). As a result, Meier *et al*. (2016) supposed that switching from 'MiSeq' to 'HiSeq' would reduce the NGS cost by a factor of 5–10. The 'HiSeq' 2500 can also be run in rapid mode, which is less cost effective but can yield current outputs of 1 Tb in 6 days (Reuter *et al*., 2015). 'NextSeq' and 'NovaSeq' provide progressive increases in throughput and therefore additional per-specimen cost reductions (Piper *et al*., 2019, Singer *et al*., 2019). Nevertheless, the increased sequencing throughput of these platforms must be balanced with diagnostic turnaround times and the possible need for logistical efforts in sample collection and processing (Chiu and Miller, 2019). As a consequence, the 'MiSeq' platform is the most recommended as it is suitable for most DNA metabarcoding/multiplexing studies when long sequences reads are needed, with respect to sequencing accuracy output and cost (Hernandez-Triana *et al*., 2017; Kerley *et al*., 2018; Liu *et al*., 2020).

2.4.2. Third-generation HTS technologies

In contrast to the short reads delivered by second-generation HTS platforms, a new generation of sequencing technologies based on single molecule sequencing has recently appeared: sometimes referred to as third-generation HTS technologies (e.g., Pacific Biosciences; RSII and Sequel instruments; and Oxford Nanopore Technologies; MinION, GridION, and PromethION instruments). These are capable of producing much longer reads with an average sequence length of >20,000 bases (Hebert *et al*., 2017; Weirather *et al*., 2017; Tedersoo *et al*., 2018; Jain *et al*., 2018).

In 2015, Pacific Biosciences (PacBio, California, USA) launched their commercial platform, the Single-molecule real-time (SMRT) (Levy and Boone, 2018). Template preparation involves ligation of single-stranded hairpin adapters onto the ends of digested DNA; synthesis occurs in

chambers, called zero-mode waveguides (ZMWs), in which a single polymerase is immobilized at the bottom of the chamber (Levene *et al*., 2003). The DNA sequence can be read in real-time from the fluorescent signals recorded in a video (Eid *et al*., 2009) as long as polymerization occurs continuously (Reuter *et al*., 2015; Levy and Boone, 2018).

Given an average raw read length of 30 kb, PacBio allows sequencing of up to 5 kb DNA fragments of satisfactory quality (Mosher *et al*., 2014; Heeger *et al*., 2018). PacBio-based metabarcoding analyses provide greater resolution than short-read second-generation HTS tools in bacteria (Singer *et al*., 2016; Wagner *et al*., 2016; Schloss *et al*., 2019) and fungi (Tedersoo *et al*., 2018), including plant pathogens (Walder *et al*., 2017). Additionally, Hebert *et al*. (2018) reported on sequencing the DNA barcode in around 10,000 arthropod specimens simultaneously in a single run. It greatly improved identification by providing more complete coverage and reduced costs 40-fold from those of Sanger sequencing (Hebert *et al*., 2018).

In PacBio instruments, the built-in circular consensus sequencing generates multiple copies of the same fragment with a highly accurate consensus (Rhoads *et al*., 2015; Hebert *et al*., 2018). It is precise (Levy and Boone, 2018), but has high initial error rates (10-15% per base) that have improved only marginally in recent years (Reuter *et al*., 2015; Tedersso *et al*., 2018). Additionally, it is currently not readily available to every laboratory due to the high cost and limited distribution of sequencing machines (Piper *et al*., 2019). PacBio sequencers are also bulky and cannot be used outside of conventional laboratory settings (Krehenwinkel *et al*., 2018).

Of the third-generation sequencing techniques, the portable MinION device (Oxford Nanopore Technologies, Oxford, U.K.) has received much attention because of its small size (that of a cell phone) and its possibility of rapid analysis at lower cost (Loit *et al*., 2019), making it affordable to governmental institutions, research laboratories and small companies (Mikheyev and Tin, 2014; Loit *et al*., 2019; Srivathsan *et al*., 2021). As their name implies, Oxford Nanopore technology uses nanopores for DNA sequence detection (Levy and Boone, 2018). Library preparation is minimal, involving fragmentation of DNA and ligation of adapters on which the entire PCR step can be skipped (Schmidt *et al*., 2017). Finally, DNA is conditioned by the addition of a motor enzyme as well as a molecular tether (Reuter *et al*., 2015). Sequencing is accomplished by measuring characteristic changes in current that are induced as the bases are threaded through the pore by a molecular motor protein (Quick *et al*., 2014).

MinION has the capacity to produce >1,000,000 sequences per day, with average read lengths of around 20,000 bases and maximum read lengths approaching 1,000,000 bases (Mardis, 2017; Jain *et al*., 2018). Additionally, Oxford Nanopore's technology may be useful for templates that are difficult to amplify, such those high in GC content or containing inverted repeats, or that otherwise pose problems for sequencing-by-synthesis platforms such as Illumina technology (Nakamura *et al*., 2011; Andrew *et al*., 2013; Srivathsan *et al*., 2021). MinION barcodes were highly accurate (~99.9%) when compared with Illumina reference barcodes (Chang *et al*., 2020). Another advantage of the MinION pipeline that it is lightweight and only requires basic molecular lab equipment available starting at $1,000 (Giordano *et al*., 2017).

2.5. Bioinformatics analysis

Following sequencing, bioinformatics analyses are needed that typically involve a pipeline that converts HTS data into molecular Operational Taxonomic Units (MOTUs) for ensuring accurate and sensitive identification (Deiner *et al*., 2017; Piper *et al*., 2019). The workflow consists of five core steps.

2.5.1. Demultiplexing and sequence quality filtering

Demultiplexing sequences is the first step to assign sequences back to their samples of origin based on index sequences or tags are incorporated into the sequencing adapters (Zhang *et al*., 2018; Zepeda-Mendoza *et al*., 2016; Tedersoo *et al*., 2018). These tags, typically 6-9 bases in length, should differ from each other by at least four bases/indels (Lundberg *et al*., 2013; Schnell *et al*., 2015; Wang *et al*., 2018) to prevent random mutations in tags or impure synthesis to erroneously switch sequences among samples. Strategies for indexing include unique dual indexing, where adapter indices at both ends of the molecule are completely unique to the sample (Schnell *et al*., 2015; Costello *et al*., 2018). However dual indexing requires a large number of tags, increasing cost (Meier *et al*., 2016; Sinha *et al*., 2017). To reduce the cost of library preparation and to generate a sufficiently large number of amplicons with a minimum number of tags for an NGS run, a combinatorial indexing approach (Meier *et al*., 2016; Wang *et al*., 2018, Yeo *et al*., 2018), using different combinations of pairs of labelled primers can be adopted.

Following demultiplexing, any other undesirable information such as PCR primer sequences and sequencing adapters are trimmed. This is, nevertheless, a rough filtering process that requires careful consideration of parameters (Piper *et al*., 2019). Increasing the strictness of quality filtering settings resulted in decreasing numbers of reads per barcode. Which affect

ulterior diversity and abundance estimates, especially samples with expected low species abundances (Arulandhu *et al*., 2017; Piper *et al*., 2019).

2.5.2. Read assembly

With the introduction of HTS, reads shrunk in size compared to Sanger sequencing but increased in number to be able to cover all the sequenced nucleic acid (called depth of coverage expressed in folds) (Miller et *al.*, 2010). These sequences need to be reassembled using their overlapping bases to reconstruct the original sequences, called contigs (Srivathsan *et al*., 2019), and the latter are arranged together to form scaffolds (Miller et *al.*, 2010). Many computational methods have been developed and are being improved to match the needs of every use case scenario, namely *de novo* assembly (Salzberg and Pop, 2008; Miller et *al.*, 2010; Shendure et *al.*, 2017) or assembly from reference (Gnerre et *al.*, 2009; Miller et *al.*, 2010).

2.5.3. Molecular Operational Taxonomic Units: clustering and quality control

Sequence similarity cluster is known as MOTUs delimitation. Sequences within an arbitrary similarity threshold (commonly 97%) are closed into representative sequences of pseudo-species, called Molecular Operational Taxonomic Units (MOTUs) (Hebert *et al*., 2003; de Kerdrel *et al*., 2020).

Several methods have been developed for molecular species delimitation such as open reference clustering algorithm: "when reads are clustered against a reference database directly, reads that do not cluster with the references are placed *de novo* into their own new clusters" (Porter *et al*., 2018). The most popular clustering method for DNA-barcoding projects is Barcode Index Numbers (BINs) (Ratnasingham & Hebert, 2013). *De novo* clustering, "when all reads are clustered amongst themselves" (Porter *et al*., 2018), include Jmotu (Jones *et al*., 2011), Vsearch (Rognes *et al*., 2016), Usearch (Edgar, 2013). Many delimitation methods based on *De novo* clustering algorithms are available; GMYC (General Mixed Yule-Coalescent model) (Pons *et al*., 2006), PTP (Poisson Tree Process) (Zhang *et al*., 2013), Neighbor-joining NJ (Saitou and Nei, 1987) ABGD (Automatic Barcode Gap Discovery) (Puillandre *et al*., 2012; Modica *et al*., 2014) and ASAP (Assemble Species by Automatic Partitioning (Puillandre *et al*., 2021).

Taxonomic decisions are based either directly on distance measures (based on an arbitrary similarity threshold, e.g., the barcode gap) (NJ, ABGD, ASAP) or on genealogy or phylogeny-based methods (GMYC, PTP), that require preliminary tree generation (DeSalle and Goldstein, 2019). Prior specification of settings is needed, which affect the performance of delimitation methods (Luo *et al*., 2018), sometimes leading to the splitting of a single species across

numerous MOTUs or the lumping of multiple species into the same MOTU: false-positive and false-negative results can be generated (Puillandre *et al*., 2012; Piper *et al*., 2019).

The MOTU quality is controlled to allow filtering of barcode loci from erroneous sequences (Liu *et al*., 2020). Chimeric sequences that occur as concatenated products of two parent sequences resulting of incompletely expanded PCR products, which act as primers for a different closely similar sequence (Potapov and Ong, 2017), are removed. Products of non-specific information resulting from non-specific amplification, such as pseudogenes and intragenomic variants, are also removed: their presence often leads to false-positive (incorrect identification of an insect as the species of concern) and false negative identifications (Piper *et al*., 2019).

2.5.4. Taxonomic assignment

Finally comes the taxonomic assignment step. To date, the most widely used approach for taxonomic classification in DNA barcoding studies has been alignment-based tools, such as BLAST, that assume that the query sequence's taxonomy will be identical or highly similar, up to a threshold as that of the most similar sequence in a reference database (Edgar, 2016; Creedy *et al*., 2020; Hsieh *et al*., 2020). For instance, Barcode of Life Data System (BOLD) (Ratnasingham and Hebert, 2007) is the widely used DNA sequence database for insects (Liu *et al*., 2020). As it is dominated by insect COI sequences from Canada (Bennett *et al*., 2019a). GenBank is a larger public database of any DNA data for any organism; however only roughly 12% of extant insect genera are thought to be represented by a COI sequence (Porter and Hajibabaei, 2018; Meiklejohn *et al*., 2019).

All these bioinformatics steps can be separately run in several free or purchased software packages. There are a variety of platforms available for biodiversity analysis, some particularly well-suited for beginners with commonly used command-line tools as well as well-documented usage examples in online forums (Porter and Hajibabaei, 2018), that remove the need for bioinformatics expertise, such as the DADA2 pipeline (Callahan *et al*., 2016), UNOISE (Edgar, 2016), QIIME (Caporaso *et al*., 2010), SCVUC COI metabarcode pipeline (Porter and Hajibabaei, 2018) and Geneious (Kearse *et al.*, 2012).

3. Conclusion

The aim of this review is to set the stage for standardized meta-barcoding approaches for large-scale biodiversity analysis so that the effects of the different chemistries, the benefits and

limitations of the different steps in a metabarcoding and/or sample multiplex studies are understood to ensure unbiased, rapid and low-cost assessments of species composition.

Many non-destructive DNA extraction methods are fast, simple, suitable for large numbers of samples, and inexpensive for preparing genomic DNA for PCR amplification (Guzmán-Larralde *et al*., 2017; Srivathsan *et al*., 2019; Suaste *et al*., 2019). Procedures that do not require spin columns or phenol chloroform chemistry, such as Chelex extraction and heat treatment chemistry, may be especially suitable for large studies. It is also important to mention that the choice of methodology will depend on the laboratory situation and funding, and of course the number and the diversity of samples to analyze.

The present study also reinforces the importance of barcode choice and primer validation for PCR amplification. In fact, each step affects the several components of diversity: richness, abundance, and beta diversity (Porter *et al*., 2019).

The choice of sequencing platform has an important impact on final results (Reuter *et al*., 2015). To date, no comprehensive benchmarking of sequencing platforms has been established for DNA metabarcoding or multiplexing studies, yet our review highlights the higher quality of sequences generated by Illumina 'Miseq', relative to its cost. Third generation HTS platforms may soon expand and supplant the current capability of DNA meta-barcoding approaches.

4. Research motivation

Globally, the major challenge in entomology is obtaining a complete inventory of all insect species (Hebert *et al*., 2016; Yeo *et al*., 2018; Seibold *et al*., 2019; Wagner, 2020). In fact, insect biodiversity exploration usually relies on trapping methods of relatively low taxonomic specificity. Consequently, researchers must often process vast and diverse collections, highlighting the urgent need for a rapid, efficient, and cost-effective DNA barcoding approach targeting arthropods. Additionally, many researchers have proposed protocols that can help with processing specimen-rich biodiversity samples at low cost (Yu *et al*., 2012; Wong *et al*., 2014; Meier *et al*., 2016). Meier *et al*. (2016) described a HTS sequencing protocol that classified MOTUs of 1015 specimens of tropical midges (Diptera: Chironomidae) at a DNA barcoding cost <1 USD per specimen. An alternative approach to species discovery is the 'reverse workflow' of Wang *et al*. (2019), using simplified DNA extraction protocols and high-throughput sequencing. Tens of thousands of specimens can be barcoded on a single Illumina lane with the total cost of a barcode being as low as $0.50 per specimen. Srivathsan *et al*. (2019) applied a workflow process using a MinION sequencing on 8700 specimens of Phoridae to

reveal diversity more comprehensively by relegating the sorting to species level to sequencing in less than a month (Srivathsan *et al*., 2019).

5. Research objectives

During our studies culminating in this present thesis, we aimed to develop a reliable, cost-effective and fast DNA meta-barcoding workflow. Our goal was to overcome the aforementioned disadvantages by targeting the three main steps in DNA barcode acquisition, DNA extraction, amplification, and sequencing. Our focus was to streamline methodologies to accomplish the following research objectives (Figure 1.1).

(1) Validate the congruence between DNA-based putative species units (MOTUs) and Morphological units (MorphOTUs), initially sorted using relevant taxonomic keys. We answer the question about the effectiveness of MOTUs vs MorphOTUs to understand the trustworthiness of MorphOTUs to evaluate the insect diversity.
(2) Show how much molecular OTUs based mini-barcodes are effective for use in Hymenoptera biodiversity assessments.
(3) Study the contribution of genetic diversity to the morphological taxonomic diversity of hundreds of Hymenoptera in the Laurentian Forest ecosystem, for example to uncover auxiliary information associated with haplotype sequences such as species crypsis.

Our long-term goal is to apply our HTS barcoding methods to cost-effectively process specimen-rich insect samples (thousands and millions). This workflow should be utilized widely due to its relative simplicity and dependence on well-established conventional laboratory procedures. It will aid in the improvement of biodiversity assessment and monitoring projects.

6. Thesis content

This thesis, entitled “High Throughput DNA barcoding to assess the diversity of Laurentian entomofaune”, includes this introductory chapter, and two research chapters, the latter two presented each as a publishable manuscript. This first chapter is a literature review of the modern methods of DNA extraction, PCR amplification and high-throughput sequencing (HTS) platforms that could expand the current capability of DNA meta-barcoding approaches, highlighting the benefits and drawbacks of each method.

The second chapter is a manuscript entitled “DNA barcoding protocol optimization for Canadian Hymenoptera” describing the DNA barcoding methods we developed for our research

and why they were chosen. We optimized two laboratory steps: DNA extraction and PCR amplification. We tested two methodological hypotheses. The first is that the QuickExtract DNA extraction method is more economic, accurate and rapid than a commonly used spin-column method. We compared its performance on the quality and the quantity of DNA that satisfy the objective to deal with Hymenoptera of different sizes and different taxonomic groups. The second hypothesis is that COI mini-barcodes behave as well as COI full length barcodes to discriminate molecular OTUs among and between Hymenoptera. We tested the performance on a variety of Hymenoptera of several published COI barcode primers, including mini- and full-length barcodes. We wanted to determine the impact of barcode length on MOTU delimitation and identification by *in silico* and *in vitro* tests. We wanted to select short barcode as locus because it is more economic and it is more appropriate for HTS sequencing. Other optimizations such as PCR annealing temperature the choice of Taq- polymerase were also evaluated.

It is important to go through this optimization step in order to validate methodological hypotheses and have a uniform protocol that is applied afterwards on the large dataset that is used in the next chapter.

The third chapter includes the optimization of the HTS-sequencing step. It is the main chapter to test the biological hypotheses of my thesis entitled "Multiplex High throughput DNA barcoding to assess Hymenoptera diversity in the Laurentian Forest –What's the congruence between morphological and molecular identification approaches?" The purpose is to respond to the third hypothesis, that MOTU-based mini-barcodes are more accurate and effective than MorphOTUs. It consists in cross-validating assigned morphological and molecular operational taxonomic units. Validation of congruence between DNA-based putative species units (MOTU) and MorphOTUs is based on OTU delimitation and OTU identification. For the OTU delimitation, the purpose was to check if the delimitation of molecular OTUs corresponds to MorphOTUs both within and between OTUs, based on the number of clusters of molecular and morphological OTU and sequence distribution within and between molecular clusters. Based on these results, we were able to reassess and possibly correct morphological OTUs, depending on their level of species-crypsis. MorphOTUs initially sorted using relevant taxonomic keys were compared with matches from GenBank (NCBI) and BOLD following a BLAST search of barcoded specimens. We wanted to understand how much we can trust MorphOTUs and how much MOTU-based mini-barcodes are effective for biodiversity assessment purposes.

We finish with a perspective in order to understand how much differences between molecular and morphological results can affect beta diversity of Hymenoptera. It will be interesting to pursue networking analysis to show further data related to haplotype sequences, such as the geographic and temporal points at which the samples were collected.

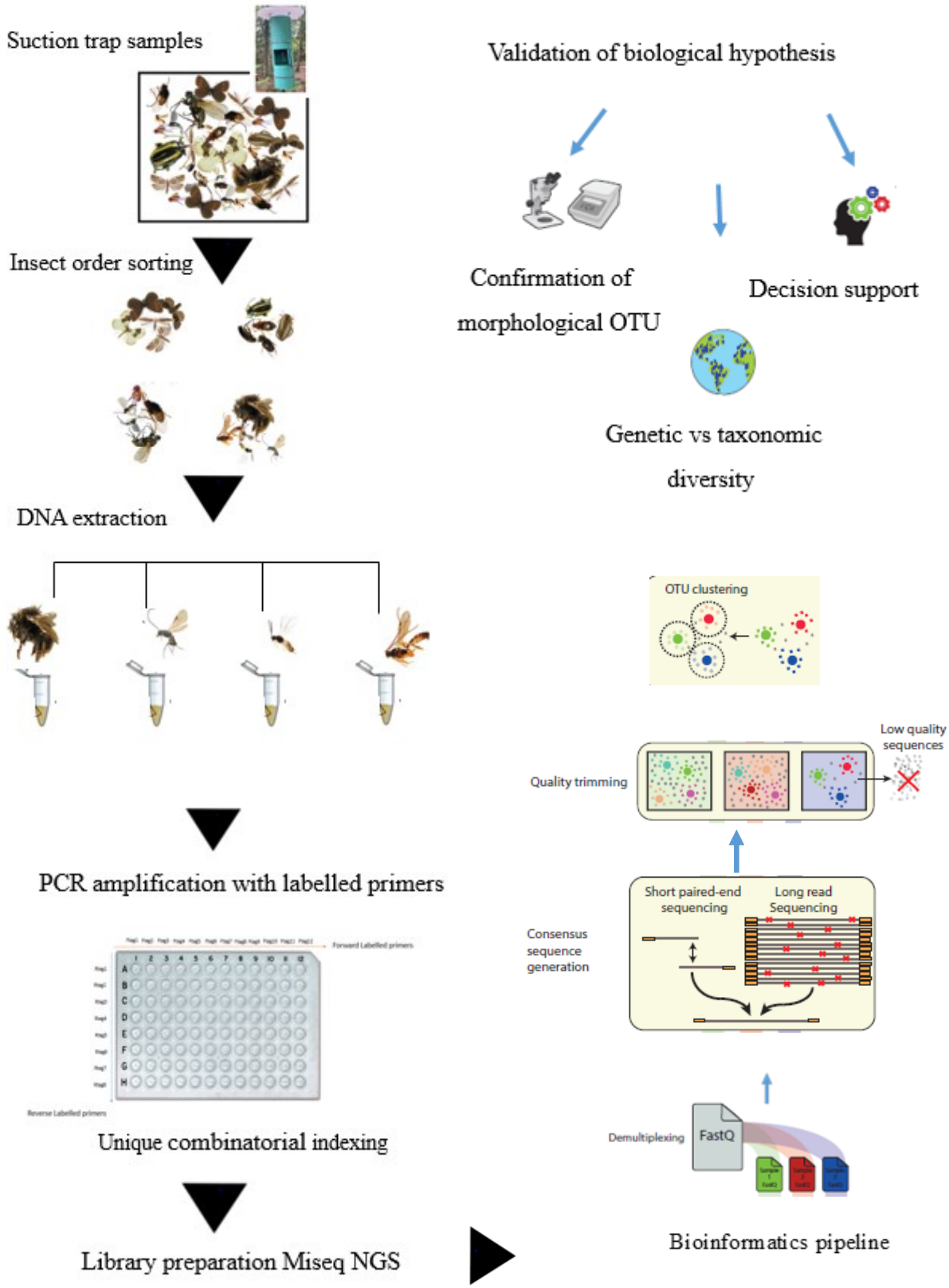


Figure 1.1. Schematization of the research workflow

Chapter 2

DNA barcoding protocol optimization for Canadian Hymenoptera

DNA barcoding protocol optimization for Canadian Hymenoptera

Malek. Kalboussi[1], Colin. Favret[1]

[1] University of Montreal, Plant Biology Research Institute (IRBV), Biology Centre, 4101rue Sherbrooke Est, Montréal, Québec, H1X 2B2, Canada.

Authors contribution

Malek. Kalboussi[1,] Conception, reaserch work and writing

Colin. Favret[1,] Conception, editing, and revision

Corresponding author. Malek kalboussi. Email: malek.kalboussi@umontreal.ca

In preparation to be submitted

Abstract

Insects are the most diverse organisms on earth with more than one million described species and millions yet to be discovered. Comprehending, discovering, and monitoring biodiversity is crucial in an era of accelerating biodiversity loss. This goal can be accomplished with the help of DNA barcoding. However, accurate, low-cost, and simple barcoding techniques are necessary to process thousands or even millions of specimens. We sought to optimize and improve barcoding of one of the most species-rich insect orders in Canada, Hymenoptera. Specifically, we compared two DNA extraction methods and evaluated the utility of mini-barcodes (208-36bp) in lieu of full-length barcodes (658bp). We compared a heat treatment-based DNA extraction method, Lucigen QuickExtract, with a commonly used spin-column method, Geneaid gSYNC, to determine the effect on DNA yield, the time required, and the cost. The heat treatment method provided a greater yield, was cheaper (0.20CAD per specimen) and faster (>25min) than the spin-column. We used three different species delimitation methods Neighbor-joining (NJ), Automatic Barcode Gap Discovery (ABGD) and Assemble Species by Automatic Partitioning (ASAP) with Hymenoptera barcode sequences *in silico* to test whether barcode length affected the number of recovered molecular operational taxonomic units (MOTUs). We found no significant difference between the number of MOTUs based on mini-barcodes and those based on the full-length barcode relative to the downloaded Barcode Index Number (BINs). We also conducted *in vitro* analysis to test the effectiveness of a set of primers and the congruence between Morphological Operational Taxonomic Unit (MorphOTUs) and Molecular Operational Taxonomic Units (MOTUs) based on mini- and full-length barcodes. We found that 313bp mini-barcodes performed similarly as full-length barcodes for specimen-level identification. The mean overlap between MorphOTUs and MOTUs was >90% for both mini- and full-length barcodes. We thus conclude that mini-barcodes, of lengths between 208 and 361bp, discriminate molecular operational taxonomic units (MOTUs) as well as full barcodes of 658bp.

Keywords : DNA extraction, COI, mini-barcodes, PCR

1. Introduction

Given the enormous diversity and abundance of insects, even minor improvements in cost-effectiveness of DNA barcoding protocols can have large benefits. This is especially true of DNA extraction and PCR amplification as we seek to increase DNA yield while minimizing contamination, degradation, and financial and time costs (Chen *et al*., 2010; Asghar *et al*., 2017; Paydar *et al*., 2018; Murthy *et al*., 2022).

In the last two decades, several improved techniques for extracting DNA from different insects have been established (Hernandez *et al*., 2012; Bahder *et al*., 2015; Giantsis *et al*., 2015; Miura *et al*., 2017; Patzold *et al*., 2020). Commercial insect DNA extraction kits are quick but expensive, yielding modest amounts of DNA (Johnson *et al*., 2001; Chen *et al*., 2010; Gupta and Preet, 2012; Giantsis *et al*., 2015; Djurhuus *et al*., 2017), and sometimes requiring expensive laboratory equipment that is not always readily available (Minas *et al*., 2011; Miura *et al*., 2017; Suaste *et al*., 2019; Liu *et al*., 2020). Some techniques using toxic solvents are harmful (Chen *et al*., 2010; Asghar *et al*., 2015); procedures necessitating maceration of the material are destructive, whereas preservation of reference specimens is essential to establish traceability between DNA and morphological traits (Martoni *et al*., 2019; Piper *et al*., 2019). Researchers and entomological laboratories need a relatively non-destructive, quick, low-cost, and high throughput DNA extraction process. Nonetheless, to our knowledge, there is no standard high-throughput genomic DNA extraction protocol for Hymenoptera from different groups.

This insect order is one of the most ecologically significant, diverse and abundant on Earth (Aguiar *et al*., 2013; Forbes *et al*., 2018; Stork *et al*., 2018) and in Canada (Bennett *et al*., 2019a). Only a small portion of their species diversity has been described (Saunderes, 2018; Bennett *et al*., 2019a). One of the causes of this lack of documentation is the insufficient molecular data reflected in the limited quantity of reference barcodes in the BOLD database (Ratnasingham and Hebert, 2007), partly due to low success rates of DNA barcoding obtained with Hymenoptera (Kaartinen *et al*., 2010; Cruaud *et al.*, 2019; Vasilita *et al*., 2022). Microhymenoptera such Mymaridae and Trichogrammatidae are tiny insects (Bennett *et al*., 2019a; Ulmer *et al*., 2021). Members of the super-families Diaprioidea and Platygastroidea are heavily sclerotized, yielding low DNA quality and quantity (Belokobyl'skij *et* al., 2019). Their biology as parasitoids depending on an insect host, make PCR amplification a challenge, which can be infected by undesirable products like host tissue and secondary endosymbionts (Kaartinen *et al*., 2010; Ulmer *et al*., 2021; Vasilita *et al*., 2022).

PCR amplification protocols can also be optimized for reliable and economic high-throughput DNA barcoding. Some DNA meta-barcoding benefits from the amplification of smaller regions of the animal barcode gene (cytochrome oxidase c subunit 1, COI). These so-called mini-barcodes successfully identify specimens (Hajibabaei *et al*., 2006; Meusnier *et al*., 2008) for surveillance (Batovska *et al*., 2017) or biodiversity assessment purposes (Meier *et al*., 2016; Wang *et al*., 2018). They are more adaptable for short-read high throughput sequencing (HTS) and can more efficiently amplify degraded DNA (Lanner *et al*., 2019; Yeo *et al*., 2020). Barcoding optimization involves the selection of primers, DNA polymerase and PCR cycling conditions (Lorenz, 2012; Elbert *et al*., 2017).

The aim of this research is to select an affordable, efficient and non-destructive DNA extraction method that does not use harmful chemical reagents or conventional enzymes. We compared the performance of a heat extraction method, QuickExtract (Lucigen, NYC, USA), with a commonly used spin-column method, gSYNC (Geneaid, New Taipei, Taiwan). We also tested the performance on a variety of Hymenoptera of several published COI barcode primers, including those for mini- and full-length barcodes. We wanted to determine the impact of barcode length on MOTUs delimitation and identification by *in silico* and *in vitro* tests. Other optimizations, such as PCR annealing temperature and the choice of Taq-polymerase were also evaluated.

2. Material and methods

2.1. Insect materials

Hymenoptera were collected from suction traps placed at the University of Montreal's Laurentian Biology Research Station (St-Hippolyte, Quebec, 45.98, -74.01) (Savage, 2001) from May to October 2021. Sampling protocols were the same as described by Favret *et al*. (2019). Insects were sorted and separated to different Morphological Operational Taxonomic Units (MorphOTUs), via a number of easily observable characters such as color, body length, leg length, wing venation, antenna shape and segment number.

2.2.DNA extraction optimization

Two DNA extraction methods were used on a total of 40 specimens ranging in size from 350 to 4000µm, from the following super-families: Chalcidoidea, Diaprioidea, Ichneumonoidea, Ceraphronoidea and Platygastroidea.

2.2.1. Lucigen QuickExtract™

Specimens were each placed individually in 20µL of QuickExtract solution, topped up as needed to completely cover the specimen. Per the manufacturer protocol, the solution with the insect was incubated at 65°C for 15min, gently vortexed, and transferred to 98°C for 2min (Figure 2.1). The intact specimen was then removed with flame-sterilized forceps and transferred to 70% ethanol for long-term storage. QuickExtract solution containing extracted DNA was then stored at −20°C until PCR amplification.

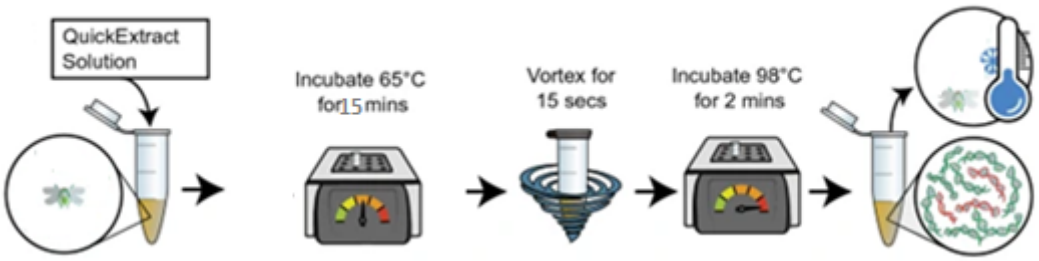


Figure 2.1. The QuickExtract DNA extraction solution (modified from Batovska *et al*., 2021)

2.2.2. Geneaid gSYNC spin-column extraction

The Geneaid gSYNC kit uses a similar chemistry and protocol as the more popular DNeasy® Blood & Tissue Kit (Qiagen, Hilden, Allemagne), but at approximately half the price. The whole insect specimen was placed in 100µL lysis buffer with 20µL of proteinase K (28,8mg/mL), and incubated for 12h at 55°C. The specimen was transferred to 70% ethanol

with sterile forceps. The DNA purification process then proceeded per the manufacturer protocol, including the addition of RNAse (50mg/mL). The purified DNA was suspended in 80µL of elution buffer.

2.2.3. Quantitative evaluation

The quality and quantity of DNA extracted from samples with the two DNA extraction methods were estimated spectrophotometrically using a Nanodrop ND-1000 spectrophotometer at 260 nm (A260) and 280 nm (A280) absorbance (Nanodrop Technologies, Wilmington, USA). DNA purity was determined using the A260/A280 ratio. Measurements were repeated three times for each sample and means used for statistical analysis.

Statistics were run with R version 4.1.1 using the "car" package (R Core Team, 2021). The DNA yield rate (log-transformed for normal distribution) and absorbance ratio (A260/A280) were evaluated using one-way ANOVA at threshold of $p<0.05$. For each response variable, homogeneity of variances (i.e., variance between factor levels is relatively equal) and residual normality were assessed with Levene's (Levene, 1960) and Shapiro-Wilk test (W).

Thirty specimens were used to study the correlation between the concentration of nucleic acid obtained with QuickExtract and the insect body length in a fixed QuickExtract solution of 50µL. The Pearson correlation method was performed by "cor.test" function in R version 4.1.1. This coefficient shows strong a linear link exists between two variables. It has a range of values from –1 to 1, with –1 indicating total negative linear correlation, 0 indicating no correlation, and +1 indicating total positive linear correlation (Bonett & Wright, 2000). Normality of the data were firstly verified by Shapiro test.

The time it takes to complete one extraction from a specimen using each of the two procedures, excluding the stages of solution preparation was calculated. The cost of one extraction was computed using the prices of the DNA extraction kits, chemical reagents, enzymes, and disposable items for each procedure (e.g., centrifuge tubes, pipette tips).

2.2.4. Qualitative evaluation

To assess DNA quality, an electrophoretogram was carried out using 1µL of the suspended DNA and directly visualized on 1% agarose gel stained with GelRed (Biotium, Fermont, USA) at 92V after 30min. DNA barcodes were amplified from each extraction. Each 20µL PCR reaction mix contained 4µL of 5× reaction buffer, 1µL of DNA template, 1.6µL of dNTPs (10 mM), 0.2µL of GoTaq® DNA Polymerase (Promega, Madison, USA), 1µL of 10µM of each

primer, mlCOIintF and modified jgHCO2198 (313bp; Leray *et al*., 2013; Meier *et al*., 2015), and 12.2µL of ultrapure water. The PCR thermal regime consisted of 35 cycles of 45s at 94˚C, 60s at 45˚C and 90s at 72˚C.

2.3. PCR amplification optimization

Our first objective was to select the appropriate taxonomic DNA barcode locus and the most adaptable primer set. Our first target locus was a mini-barcode, 313bp within the 658bp of the full-length barcode marker. Mini-barcode amplicons are easier to obtain when the DNA in the sample is degraded and they can be sequenced at lower cost on short-read sequencing platforms. However, as the shorter sequences contain less diagnostic information, they may result in less accurate taxonomic identifications. We sought to ascertain how less accurate they would be and whether the shorter sequences would be fit-for-purpose nonetheless. *In silico* and then *in vitro*, we evaluated the taxonomic resolving power of several mini-barcodes, as compared to full-length barcodes, based on several available primer sets.

2.3.1. *In silico* analysis: Species delimitation

We compared the performance *in silico* of 658bp of COI barcode vs mini-barcodes using three different molecular operational taxonomic unit (MOTU) clustering methods: two automatic barcode gap identification methods called Automatic Barcode Gap Discovery (ABGD) (Puillandre *et al*., 2012) and Assemble Species by Automatic Partitioning (ASAP) (Puillandre *et al*., 2021), and Neighbor-joining (NJ); a frequently used dendrogram-based method for species delimitation (Hong *et al*., 2021). We evaluated if full-length barcodes and mini-barcodes both recovered a similar number of MOTUs in the three different clustering methods.

2.3.1.1. Dataset selection

The dataset, download from BOLD, contained 1039 DNA barcodes of Canadian Hymenoptera, with a length of 660bp, belonging to nine different Barcode Index Numbers (BINs), BOLD's equivalent of MOTUs. Six of the nine MOTUs were of the same family and subfamily (Braconidae: Microgastrinae) (Table 2.1).

Table 2.1. Summary of the DNA barcode sequences downloaded from BOLD and used in clustering analysis

Barcode Index Numbers BINs	Sequence Count	Species count	(Family: subfamily)	Molecular OTUs number
BOLD: AAA4780	40	1: *Glyptapanteles*	(Braconidae: Microgastrinae)	MOTU1
BOLD: AAA6373	147 1	1: *Apanteles* 1: Unidentified	(Braconidae: Microgastrinae)	MOTU2
BOLD: AAA6712	86	1: *Lissonota*	(Ichneumonidae: Atrophini)	MOTU3
BOLD: AAA7886	105	1: *Microgaster*	(Braconidae: Microgastrinae)	MOTU4
BOLD: AAA8055	106	1: *Cotesia*	(Braconidae: Microgastrinae)	MOTU5
BOLD: AAB0136	66 5	1: *Glypta* 1: Unidentified	(Braconidae: Banchinae)	MOTU6
BOLD: AAB0186	7	1: *Dolicogaster*	(Braconidae: Microgastrinae)	MOTU7
BOLD: AAA1841	174	1: *Myrmica*	(Formicidae: Myrmicinae)	MOTU8
BOLD: AAA3764	302	1: Dolichogenidea	(Braconidae: Microgastrinae)	MOTU9

2.3.1.2.Alignment and excising *in silico* mini-barcodes

We identified four mini-barcodes with published primers within the full-length DNA barcode. The four pairs of mini-barcode primers (Table 2.2) were aligned to a braconid 658-COI gene with ClustalW in Geneious 9 software (Biomatters Ltd, Auckland, New Zealand) (Kearse *et al.*, 2012) in order to identify the precise position of the mini-barcodes within the full-length barcode. They were 208, 307, 313 and 361bp long (Table 2.2).

2.3.1.3.Species delimitation methods

Neighbor joining's algorithm uses the data and generates a distance matrix between candidates and then estimates a tree that matches this information (Saitou and Nei, 1987). ABGD and ASAP are methods based on pairwise genetic distances to delimit species. ABGD, "Automatic Barcode Gap Discovery", compares the gap existing between the range of intra and interspecific sequence distances to automatically find the level of sequence divergence where the barcode gap is located (Puillandre *et al*., 2012). ASAP, "Assemble Species by Automatic Partitioning", uses the same principle as ABGD, however it provides a score for each defined partition to overcome the challenge of the prior method (Puillandre *et al*., 2021).

All three methods take as input a Fasta-formatted sequence alignment file; alignments were performed with the default options of the ClustalW tool in Geneious (Kearse *et al.*, 2012). The input files were then used to compute a matrix of pairwise distances using the Kimura two parameter model, the method that best fitted the data than Jukes-Cantor (JC69) method. We applied ABGD and ASAP (Kimura 1980) through their web interfaces, https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html and https://bioinfo.mnhn.fr/abi/public/asap, respectively, for ABGD and ASAP. We executed ABGD with the following settings: (P)min=0.001, (P)max=0.1, and 20 steps, with gap width=1. We considered the ABGD clusters at priors 0.02<(P)<0.1. For ASAP, we used a recursive split probability of 0.01 and reported the partition with the best asap-score (1). Tree construction using neighbor joining's algorithm was performed with Geneious 9 software (Kearse *et al.*, 2012) with the default options except for the number of bootstrap replicates that was set to 1000.

2.3.2. *In vitro* analysis: species discrimination and identification

We compared the identification based on mini-barcodes and full barcodes obtained with fresh material in the laboratory. The MorphOTU identities based on morphological examination of specimens were compared to barcodes identities from specimens belonging to each MorphOTU, following an NCBI BLAST search of mini- and full-length barcodes in both GenBank and BOLD databases. We first selected the primer sets to amplify mini- and full-length barcodes as well as the Taq polymerase.

2.3.2.1.Primer and taq polymerase selection

We evaluated the performance of four primer sets targeting COI for 36 specimens. Using three different taq polymerases: TaKaRa Ex Taq™ (Thermo Fisher Scientific, Waltham, MA), Phire

Green Hot Start II DNA Polymerase (Thermo Fisher Scientific, Waltham, MA) and Go taq. PCR mixtures (Promega, Madison, USA), was the same as the previously described in the DNA quality evaluation section (2.2.4). Four µL of 5× Go taq and Phire Green Hot Start II DNA Polymerase, or 2µL of 10× TaKaRa ExTaq rTaq reaction Buffer were used.

We used the three annealing temperatures published in the corresponding articles of each primer set: 46°C (Gibson *et al*., 2014), 47°C (Leray *et al*., 2013; Meier *et al*., 2016) and 45°C (Folmer *et al*., 1994), respectively, for ArF10/ ArR5, m1COlintF/Modified jgHCO2198, LCO1490 /HCO2198 and Lep-f1/LepR1 (Théry *et al*., 2017) (Table 2.2). Initial denaturation at 98°C for 1min was followed by 35 cycles of denaturation at 94°C for 5s, extension at 72°C for 1min:30, and final extension at 72°C for 2min. For each PCR, we verified successful amplification and measured the strength of the amplification product via agarose gel electrophoresis. One-way ANOVA analyses were conducted in R to compare the polymerase success rate.

Table 2.2. Mini- and full-length barcode primer sequences and annealing conditions

Mini-barcode Length (bases)	**Barcode Primer Sequence (5' → 3')**	**Annealing conditions**	**Reference**
658	F: HCO2198 : TAAACTTCAGGGTGACCAAAAAATCA	45°C for 1min	(Folmer *et al*., 1994)
	R: LCO1490 GGTCAACAAATCATAAAGATATTGG		
658	F :LepF1 ATTCAACCAATCATAAAGATATTGG	48°C for 1min	(Favret *et al*., 2005), (Théry *et al*., 2017)
	R: LepR1 TAAACTTCTGGATGTCCAAAAAATCA		
307	F: ArF10: CCWGATATAKCITWYCCICG	46°C for 1min	(Gibson *et al*., 2014)
	R: ArR5: GCICCRGAYATRGCITTYCCACG		
313	F: m1COlintF: GGWACWGGWTGAACWGTWTAYCCYCC	47°C for 1min	

	R: Modified jgHCO2198: TANACYTCNGGRTGNCCRAARAAYCA		(Leray *et al*., 2013), (Geller *et al*., 2013)
361	F: ArF10: CCWGATATAKCITWYCCICG	47°C for 1min	(Leray *et al*., 2013), (Geller *et al*., 2013)
	R: Modified jgHCO2198: TANACYTCNGGRTGNCCRAARAAYCA		
208	F: m1COlintF: GGWACWGGWTGAACWGTWTAYCCYCC	46°C for 1min	(Gibson *et al*., 2014), (Leray *et al*., 2013)
	R: ArR5: 5′-GCICCRGAYATRGCITTYCCACG		

2.3.2.2.PCR amplification and DNA sequencing

In order to identify the accuracy of the mini-barcodes within the full-length barcode and validate the efficiency of the chosen primer pairs, we tested amplification success rates for 46 specimens belonging to 23 MorphOTUs. The same genomic DNA was used for each specimen to amplify the full-length barcode and a mini-barcode of 313-bp length in two separate PCRs using different primers: HCO2198 and LCO1490 (658-bp; Folmer *et al*., 1994), mlCOIintF and modified jgHCO2198 (313-bp; Leray *et al*., 2013). The PCR reagents were 10µL of Phire Green Hot Start II PCR Master (Thermo Fisher Scientific, Waltham, MA), 2µL of 1 mg/mL BSA, 1µL each of 10µM primers, and 1µL of DNA. The PCR conditions were 40s initial denaturation at 98°C followed by 35 cycles of denaturation at 98°C for 5sec, annealing at 47°C for 50sec, extension at 72°C for 15sec, followed by final extension of 72°C at 1min. PCR products were Sanger sequenced at Genome Quebec (Canada, Montreal). The quality control, trimming, consensus generation and alignment were performed using Geneious 9 software (Kearse *et al.*, 2012).

Representative sequences were taxonomically assigned based on GenBank (NCBI) and BOLD (Ratnasingham and Hebert 2007) databases. Barcodes that could not be matched to a family, genus or species with high confidence (≥95% identity at 100% query cover) were presumed to be contaminants and thus omitted.

2.4. Direct PCR optimization

Direct polymerase chain reactions (dPCR) were performed to amplify DNA barcodes from 20 specimens raging in size from 375 to 9010µm, from following the super-families: Chalcidoidea, Diaprioidea, Ichneumonoidea, Ceraphronoidea and Platygastroidea. Amplification procedures were adapted from Wong *et al*. (2014), Meier *et al*. (2016), and Wang *et al*. (2018), and optimized for Hymenoptera as follows. We adjusted the tissue amount for dPCR according to the Hymenoptera size. One leg or two legs were excised from large (>9000µm) and medium specimens (3000-9000 µm), respectively. The whole specimen was used to for small and very small insects (<1500-3000µm). PCR conditions were the same as previously described for m1COlintF and Modified jgHCO2198 primers. We used 10µL of Phire Tissue Direct PCR Master Mix (Thermo Fisher Scientific, Waltham, MA) instead of Phire Green Hot Start II PCR Master. After PCR, the body parts inside the PCR wells were removed, preserved as vouchers in 70% ethanol and stored at –20 °C. The amplification results were checked with electrophoresis on a 1% agarose gel stained with GelRed.

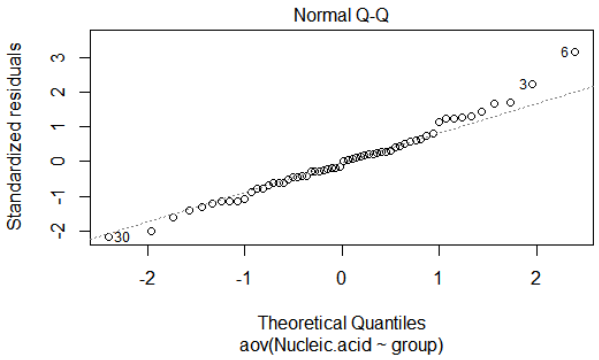
3. Results

3.1.DNA extraction optimization

3.1.1. Quantitative evaluation

3.1.1.1. Data normality verification

The Shapiro Wilk-Test (W) indicated that the nucleic acid concentration and absorbance ratio did not deviate significantly from normal distribution at 5% threshold (W=0.96318, p=0.06741) (W=0.98175, p=0.5068). Data approximately follow a straight line of residuals normally distributed (Figure 2.2). Additionally, Levene's test confirmed the homogeneity of variance within the two groups (Levene, 1960) with no significant difference between them (p=0.324 and p=0.065), respectively, for the nucleic acid concentration and the absorbance ratio.



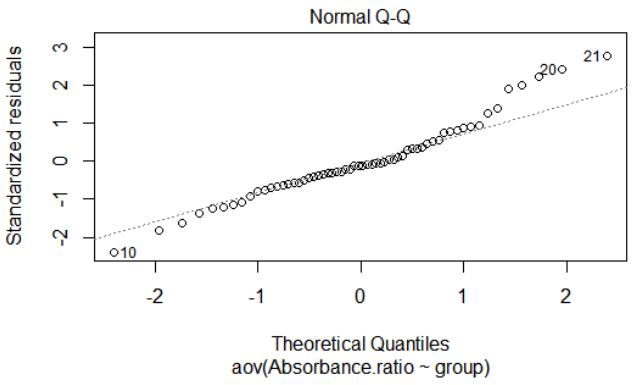


Figure 2.2. Normality plot of our data for the Nucleic acid concentration (left) and Absorbance ration variable (right)

3.1.1.2. DNA quantity

DNA yield was greater with the QuickExtract method than the spin-column. The nucleic acid concentration varied from 2.5 to 84.6ng/µL with an average of 31.26±0.49ng/µL based on the three repetitions. The nucleic acid concentration obtained with gSYNC kit ranged from 1.5 to 12.46ng/µL with an averageof 6.4±0.7ng/µL in 80µL of elution buffer. Statistically, DNA nucleic acid was significantly higher ($p<2e-16$) with Quick Extract than that obtained by the kit used at the 5% level of significance (Figure 2.3).

Taking into account the variable volume of the QuickExtract solution, we calculated the DNA amount. The later varied from 200 to 1692ng with an average of 937,8±12,3ng, statistically higher ($p<0.05$) than DNA amount obtained with gSYNC kit between 120 and 977ng with a median of 512±22.3ng.

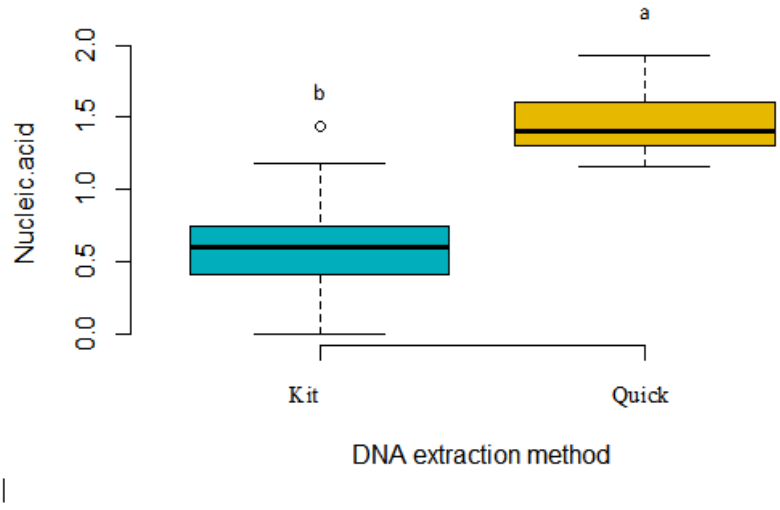


Figure 2.3. Box-plot of absorbance ratio obtained with QuickExtract and the gSYNC kit (mean plot of absorbance ratio followed by the same letters are not significantly different according to the Tukey test, $p<0.05$).

3.1.1.3. The absorbance ratio A260/A280

The absorbance ratio ranged from 0.89 to 2.1 for QuickExtract (unlike with the gSYNC kit we did not include an RNA degradation step) and 1.15 to 2.13 with gSYNC kit, with an average of 1.70±0.05 and 1.79±0.07 respectively for QuickExtract and gSYNC kit. Statistically, the

analysis of variance did not show any significant effect of the DNA extraction method used at 5% level for the absorbance ratio A260/A280 (p=0.091) (Figure 2.4).

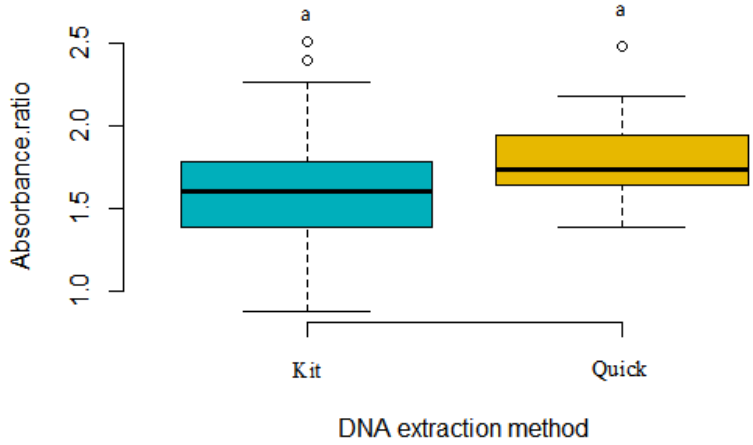


Figure 2.4. Box-plot of absorbance ratio obtained with QuickExtract and gSYNC kit (mean plot of absorbance ratio followed by the same letters are not significantly different according to the Tukey test, $p<0.05$).

3.1.1.4.DNA yield and specimen size

The Pearson correlation assigned a positive value of 0.56. Since this probability is low (p=6,798 e-05) (below the significance threshold of $p<0.05$), H0, "null hypothesis" which is the probability of having such a high correlation in a sample if the correlation in the population is zero, can be rejected. As a result, the correlation between specimen size and nucleic acid concentration is concluded to be significantly positive (Boslaugh and Watters, 2007) (Table 3). However, the Pearson coefficient is less than 1, so the increase of concentration of nucleic acid as the specimen body length increases was not constant (Berman, 2016). Such unstable correlation was observed with large Hymenoptera, especially Ichneumonoidea, which lead to low nucleic acid concentration, and very small insects like Trichogrammatidae and Mymaridae provided high nucleic acid concentrations.

According to QuickExtract method, large specimens (>9000µm) showed an absorbance ratio lower than 1.9 suggesting protein contamination (Chen *et al.*, 2010), or higher than 2, suggesting RNA contamination (Wang *et al.*, 2019). We observed that the smaller the specimen, the higher the DNA quality was (Table 2.3).

Table 2.3. DNA absorbance ratio and yield rate of DNA extracted by QuickExtract method according to specimen body-size

DNA quality / **Specimen body-size**	**Concentration of nucleic acid (NA) (ng/µL)**	**DNA absorbance ration (AR) A260/A280**
Very small specimens (< 1500µm)	10< (NA) <30ng/µL	1.76< (AR) <1,89
Small-sized specimens (1500-3000µm)	22< (NA) <50ng/µL	1.77< (AR) <1,95
Medium-sized specimen (3000-9000µm)	45< (NA) <85ng/µL	1.69< (AR) <1,79
Large-sized specimens (> 9000µm)	90<(NA) <220ng/µL	0.89< (AR) <2.1

3.1.1.5.Time and cost

By omitting the precipitation step and minimizing the incubation time, the QuickExtract method was much faster than gSYNC. The former could be completed in less than 20min (Table 2.4). The gSYNC kit was more time consuming, taking 14 to 26 h. It was also more expensive. The QuickExtract solution was the only reagent needed whereas the gSYNC method required multiple transfers of various solutions and buffers.

3.1.2. Qualitative evaluation

Electrophoretic smears were brighter with the QuickExtract than those of spin-column extraction. However, smear tails likely caused by DNA degradation or RNA contamination (Wang *et al*., 2019) were observed for 20% of samples with QuickExtract. Using gSYNC we obtained lower quality (per the spectrophotometric absorbance ratio) but smear tails were absent presumably because we used the RNase A during the extraction process.

The PCR amplification rate was 93% for QuickExtract DNA extractions and 66% for spin-column extractions. DNA bands obtained with QuickExtract PCRs were brighter than those obtained using gSYNC (Figure 2.5).

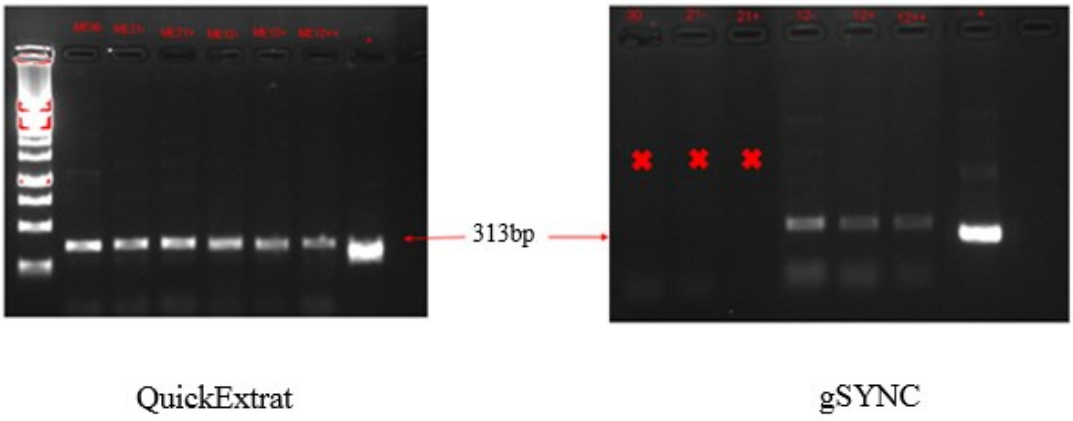


Figure 2.5. PCR amplification pattern of COI barcode for testing the DNA quality extracted specimens.

The QuickExtract method preserved the specimen's features after extraction (Figure 2.6), even for small insects. The gSYNC method required a perforation of the abdomen to improve DNA yield which was destructive for small specimens and making morphological identification nearly impossible.

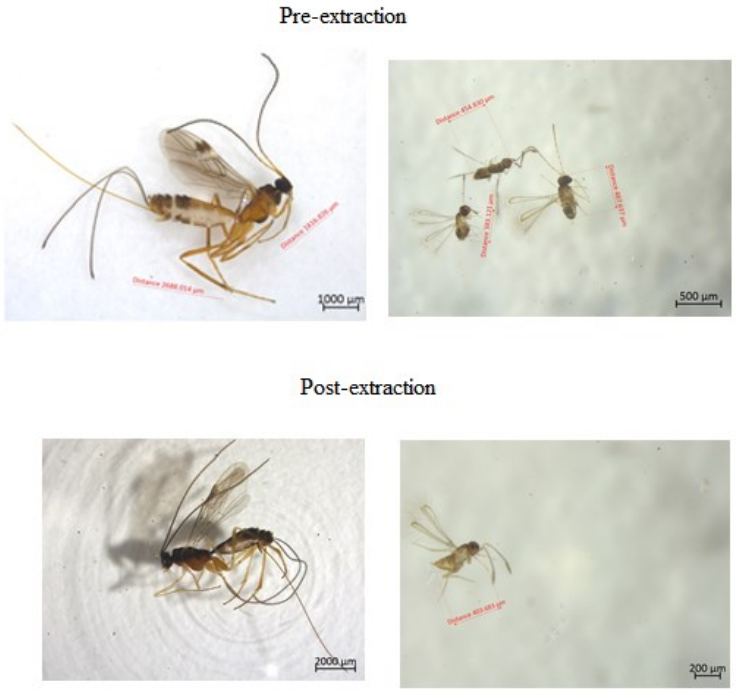


Figure 2.6. Hymenoptera from a trap sample before and after non-destructive DNA extraction method QuickExtract

Table 2.4. Estimated cost and time required per extraction by using QuickExtract and gSYNC used methods

Variable	DNA extraction solution	
	QuickExtract	gSYNC
Time(min)	<20min	25min+12h incubation
Cost per specimen	<0.5CAD	>3CAD
Specimen features	Non-destructive	Partially destructive
Simplicity	+++	+
Automatisation	Relatively easier	More complex

3.2.PCR amplification optimization

3.2.1. *In silico* analysis: Species delimitation

The number of predicted BINs provided by BOLD is only a proxy to assess the performance of mini- and full-length barcodes (controlled scenario). The number of MOTUs reported by each delimitation method varied between nine and 11 according to different lengths barcodes (Table 2.5).

3.2.1.1. Neighbor joining output

Neighbor-joining recovered nine MOTUs subjectively-recognized based on bootstrap values >90% and length of branches. Based on the full-length barcode dendrogram, NJ tended to consider the six groups from the same subfamily as most closely-related (Table 1). *Microgaster* (Braconidae: Microgastrinae), *Glyptapanteles* (Braconidae: Microgastrinae) and *Cotesia* (Braconidae: Microgastrinae) were grouped together, whereas Dolichogenidea (Braconidae: Microgastrinae) and *Apanteles* (Braconidae: Microgastrinae) were recovered as sister groups. *Myrmica* (Formicidae: Myrmicinae) was the most phylogenetically distant group, but it was closer to *Lissonota* (Ichneumonidae: Atrophini) and *Glypta* (Braconidae: Banchinae), which were sister groups to the other MOTUs. The 313 and 307bp mini-barcodes generated the same results as the full-length barcode, with the same MOTUs numbers and relationships. The 361 and 208- mini-barcodes tended to bring the sister group Dolichogenidea and *Apanteles* closer to *Lissonota* and *Glypta.*

3.2.1.2. ABGD output

ABGD recovered nine and ten partitions for the mini- and full-length barcodes respectively (Table 2.5). Only initial taxon partitions were considered (Figure S1). Recursion partitions that divide the initial into secondary partitions (Puillandre *et al*., 2012) were omitted, because they over split MOTUs comparing to the first one.

The number of MOTUs retained was inversally propotional to the preselected P-distance (0,1<(P)<0.001). A number of MOTUs ranged from 9 to 124 with the full-length barcode and from 4 to 361 with the mini-barcodes was reported (Figure S1). The lower range was when the intraspecific P-distances (P)=0.1, whereas the upper range was when (P)=0.001. The more stable partition provided by ABGD (generally when 0.02<(P)<0.05) included nine MOTUs for all mini-barcodes as well as full length barcode (Figure S1). The same sequences were assigned to the same MOTUs with all sequence lengths (Table S1).

3.2.1.3. ASAP output

ASAP analyses contained a variable number of MOTUs, mostly overlapping with ABGD, with prior distances (P) between 0.04 and 0.1. It generated the same result for 361bp and 307bp mini-barcodes and the full-length barcode, often with the 1st or the second best ASAP score, the same number of MOTUs (Figure S1) and the same sequences among MOTUs (Table S1). For 313 and 208bp mini-barcodes, the best ASAP score yielded 11 clusters (Figure S1). ASAP assigned three and two clusters for the MOTU9 respectively for 313 and 208 mini-barcodes: *Dolichogenidea*, with 300 sequences. It partitioned the two other sequences into two additional MOTUs (10 and 11). For the 208bp mini-barcode, ASAP generated a supplementary cluster with one sequence of *Dolichogenidea* (Table S1).

Table 2.5. Number of MOTUs recovered with four mini-barcodes and the full-length barcode using three clustering algorithms.

Clustering method / Barcode length	Number of MOTUs		
	ABGD	ASAP	NJ
Full length 658bp	9	9	9
307bp	9	9	9
313bp	9	11	9
208bp	9	10	9
361bp	9	9	9

3.2.2. *In vivo* analysis: species discrimination and identification

3.2.2.1. Primer and taq polymerase selection

With i*n vitro* PCR runs, the primer pair m1COlintF/Modified jgHCO2198 was the most consistently successful to amplify 313-mini-barcodes, with a success rate of 96.42% The other primer pairs performed relatively poorly, with 53,57% success for LCO1490/HCO2198, 51% for Lep-F1/Lep-R1 amplifying the 658 full-length barcode, and no successful amplification with ArF10/ArR5 and the 307 bp mini-barcode.

Ex-Taq and Phire performed better than Go taq, with 94% and 92% success rates for the former two and 74% for the latter. Phire yielded approximately similar high success rates as the much more expensive Ex-Taq (twice the cost) with no significant difference ($F=0.511$, $p=0.477$) and with a volume half that of Ex-taq.

3.2.2.2.PCR amplification and sequencing analysis

Having selected Phire as our polymerase of choice, we recovered amplification success rates of 96% for mini-barcodes and 53% for full length barcodes. Sanger sequencing success rates were similar: 71% for 313-mini-barcodes and 69% for full-length barcodes.

Sequencing results and their subsequent comparison with available GenBank and BOLD databases allowed the recognition and identification of the same number of MorphOTUs tested

for mini- and full-length barcodes (Table S2). We found similar BLAST identification rates for mini-barcodes and full-length barcodes for all of the *in vitro* data. The majority of taxonomic identification of Hymenoptera was restricted to the family level with sequence identity scores ranging from 95% to 100% for the consensus sequences for most of the MOTUs. In most of the cases, BOLD results were consistent with GenBank results except for one sample with a low identity score (MOTU29).

When comparing the suitability of full-length and mini-barcodes for identification purposes, we found that full-length barcodes have the highest identification percentage in databases.

3.3. Direct PCR

Direct PCR provided low success rates for our Hymenoptera overall (20%) but was promising for the smallest specimens (<1500-3000µm), with a success rate of 90%, mainly Chalcidoidea. The removal of legs was obviously partially destructive for medium and large Hymenoptera, but the presumed availability of mitochondria-rich muscle tissues through the legs did not increase the direct PCR success rate.

4. Discussion

4.1. DNA extraction

In molecular entomology, some insect samples, due to their size, their morphological features such as the high degree of sclerotization, and their biology (Kaartinen *et al*., 2010; Ulmer *et al*., 2021; Vasilita *et al*., 2022) could not be safely and easily processed with commonly used non-destructive extraction methods (Wong *et al*., 2014; Wang *et al*., 2018). Yet DNA extraction is the main step affecting the continuity of a DNA barcoding workflow (Guzmán-Larralde *et al*., 2017; Paydar *et al*., 2018). We obtained excellent results across a range of Hymenoptera using the QuickExtract method for our DNA barcoding process. These results were previously validated by Srivathsan *et al*. (2019) and Batovska *et al*. (2021), providing the assurance of morphologically intact specimens. The amplification success rate documented that the extracted DNA by QuickExtract was of high quality; sharper and more intense electrophoresis bands were obtained with PCR products originating from QuickExtract extractions.

The DNA quantity was higher with QuickExtract than gSYNC. Compared to other insects, the amount of DNA from Hymenoptera wasn't high compared to other insect groups, it ranged between (200-1692ng) with QuickExtract and (120-977ng) with gSYNC. De la Cruz-Ramos *et al*. (2019) recorded an amount of DNA between 5848–9188ng for mosquito. For mealybugs, the DNA concentrations ranged between 1254 and 11219ng depending on the DNA extraction

method used (Wang *et al*., 2019). For beetles, Murthy *et al*. (2022) obtained nucleic acid concentration varied between 1160 and 4820ng with different methods. It can be due to the little size of specimens, like mymarommatoids (Bennett *et al*., 2019b) or the high sclerotization of specimens of other super-families such as Diaprioidea and Platygastroidea (Belokobyl'skij *et al*., 2019).

The QuickExtract method provided high amplification success rates with all the superfamilies, However, the gSYNC extraction failed to yield PCR bands with specimens of Ceraphronoidea and Chalcidoidea known to be difficult to barcode (Ulmer *et al*., 2021; Vasilita *et al*., 2022).
20% of QuickExtract extractions exhibited smear tails likely caused by DNA degradation. It is likely that the omission of various purification steps with the QuickExtract method failed to remove some contaminants. However, no significant effect was observed on DNA purity between QuickExtract and gSYNC kit. The 20% observed smears could be just as present in both methods but simply more visible in the brighter gels obtained for QuickExtract. In these cases, additional tests with more standardized conditions should be realized.

Using the QuickExtract method, we report a positive correlation between nucleic acid concentration and Hymenoptera body size. The correlation is unstable because some large sized Hymenoptera are very sclerotized forbidding a high DNA yield. It was observed with some large Ichneumonoidea and Diaprioidea. Other very small insects like Trichogrammatidae and Mymaridae are lightly sclerotized, leading to higher nucleic acid concentration (Carew *et al*., 2018). Similar results were also reported by Batovska *et al*. (2021), admitting that QuickExtract works better with soft-bodied taxa than with those with higher levels of sclerotization.
We also report that specimen body length affects the 260/280 absorbance ratio. Larger specimens provided an absorbance ratio exceeding 1.8, probably due to the presence of RNA and other contaminants such as proteins and lipids (de la Cruz-Ramos *et al*., 2019; Wang *et al*., 2019) (Table 3).

Consequently, it may be beneficial to apply a longer incubation period for sclerotized specimens to release additional DNA (Asghar *et al*., 2015; Carew *et al*., 2019). While this would increase the time taken to conduct DNA extractions, it would likely be better than dissecting specimens to give enzyme access to tissues.

The cost of reagents with QuickExtract was approximately one sixth that of the spin column method and took approximately one-thirtieth of the time, taking into account the incubation step.

According to these results, the comparison in the laboratory to the spin-column gSYNC and other DNA extraction methods in the literature (Chen *et al*., 2010; Asghar *et al*., 2015; Bahder *et al*., 2015; Miura *et al*., 2017), we will use QuickExtract for our future high throughput Hymenoptera DNA-barcoding.

4.2.PCR amplification

Mini-barcodes provided a higher amplification success rates than full-length barcodes. Other researchers also reported success with mini-barcodes. Hajibabaei *et al*. (2006) demonstrated the effectiveness of mini-barcodes to amplify degraded DNA: 90% PCR success against 50% for full barcodes. Guzmán-Larralde *et al*. (2017) used mini-barcodes to sequence DNA from museum samples up to 23 years old. Meier *et al*. (2016) recorded high success rates, exceeding 80%, in amplifying tropical midge mini-barcodes (Diptera: Chironomidae). Wang *et al*. (2018) mini-barcoded 4032 ants for a total lab cost of <0.50CAD per specimen.

In *silico* analysis revealed the potential of mini-barcodes to delimit MOTUs just as well as the full-length barcode. We found that the performance of COI mini-barcodes did not differ substantially from that of full-length barcodes using ABGD and NJ clustering methods. We found the same number of MOTUs with the same composition of species for four sets of mini-barcodes as with the full-length barcode (nine MOTUs).

Unfortunately, the lack of a scoring system in ABGD and NJ makes the choice of the relevant partition less straightforward than with ASAP. Future work may seek to combine and compare a variety of clustering approaches to discover variations between partitions and clusters.

The ASAP results did not differ significantly from those of ABGD or NJ ($p<0.05$), but they behaved erratically for the 313 and 208-base mini-barcodes specifically for Dolichogenidea: We found that they underperformed the other barcodes under the lowest prior (P)=0.04. This incongruent result can be explained by two hypotheses. The mean intra-BIN variability between specimens of *Dolichogenidea* sp was 0.0132, higher than that of *Glyptapanteles* specimens, for example, which was 0.0057. As a result, ASAP considers the intraspecific genetic diversity within MOTU 9, *Dolichogenidea*, to be too low to attribute all sequences to the same cluster. Secondly, these results can mean that these mini-barcodes were not really informative to correctly assign sequences to the correct cluster, especially as it was not the same case when the full barcode was used. This may be a significant drawback of the ASAP method (Puillandre *et al*., 2021), relative to the sensitivity of the other delimitation methods: we did not encounter this problem with ABGD and NJ using the same mini-barcodes.

Based on *in vitro* analysis, the primer pair m1ColintF/Modified jgHCO2198 provided the best amplification success rate with the 313bp mini-barcode. This primer pair has provided high amplification success rates in other Hymenoptera studies: >80% for Formicidae (Wong *et al*., 2014; Wang *et al*., 2018) and other insect faunas (Leray *et al*., 2013; Yeo *et al*., 2018; Morinière *et al*., 2019; Srivathsan *et al*., 2019). Despite the promising *in silico* performance of the 307bp mini-barcode primer pair ArF10/ArR5, it failed to amplify *in vitro*. Our PCR optimization step included the primers, annealing temperature and the taq polymerase optimization in order to increase the PCR Successfulness at lower price.

In vitro analysis revealed the same identification based on 313 mini-barcode as the full-length barcode for our MOTUs in GenBank and BOLD databases. These results were approved with previous findings (Meusnier *et al*., 2008; Yeo *et al*., 2020).

In silico and *in vitro* analyses of Hymenoptera barcodes revealed the ability of mini-barcodes to discriminate among MOTUs just as well as full-length barcodes. Our aim to use mini-barcodes in our research is thus supported. Our results were congruent with Hajibabaei *et al*. (2006) who showed a high congruence with the full-length barcode when species are delimited based on mini-barcodes. In their *in silico* testing, Meusnier *et al*. (2008) discovered that mini-barcodes and full-length barcodes both had comparable BLAST identification rates. Hernandez *et al*. (2012) amplified a shorter fragment (between 200-400 bp) of COI for revealing diversity of Coleoptera. Yeo *et al*. (2020) recently conducted an extensive analysis on >5000 species, including 15,347 sequences of Hymenoptera, to understand if mini-barcodes with different lengths were able to provide similar identifications and species delimitations as 658-bp barcodes. Their results demonstrated no significant difference in performance for species delimitation and identification between full-length and mini-barcodes as long as they are of moderate length (>200-400bp). Only very short mini-barcodes (<200bp) perform poorly, especially when they are located near the 5 ’end of the Folmer region.

In contrast, Yu and You (2010) conceded that mini-barcodes may be less accurate than full-length barcodes for species identification. In addition, Sultana *et al*. (2018) found that when the barcodes are too short (150bp), mini-barcodes had a decreased capacity to distinguish species.

Here, we demonstrate that 313bp mini-barcodes are just as dependable as full-length barcodes for sorting specimens into potential MOTUs, approving our second hypothesis.

It is also crucial to mention that there are not enough data to test the performance of mini-barcodes against morphological delimitation. The goal of this chapter was to test their performance with that of full-length barcodes in order to select the barcode length for our PCR amplification step. Testing the effectiveness of DNA mini-barcodes against morphological identification, especially delimitation, is the main objective of the following chapter.

Although direct PCR was cheaper and quicker because it skips the DNA extraction step, we cannot adopt it due to a low PCR success rate. We suspect this high failure rate is due to the heavy sclerotization of many Hymenoptera, impeding access to genetic material by the PCR reagents. These results were also found by Wong *et al*. (2014) and Wang *et al*. (2018) during Direct PCR amplification on Formicidae. Direct PCR was also partially or wholly destructive and thus not in line with our objectives.

In conclusion, based on the results presented here, for our future research we adopted QuickExtract as DNA extraction method, we amplified mini-barcodes of 313bp of COI with m1COlintF and Modified jgHCO2198 primers and used the Phire DNA polymerase. This work is described in the followed chapter.

5. Supplementary material

ASAP

A B C D E

Nb of species	asap-score		P-val (rank)	W (rank)	Treshold dist.	Text
9	2.00	●	1.00e-05 (3)	5.55e-05 (1)	0.054810	list csv
4	4.50	●	1.00e-05 (1)	5.18e-07 (8)	0.128449	list csv
* 12	6.50	●	1.81e-01 (8)	1.36e-06 (5)	0.014376	list csv
9	9.50	●	1.00e-05 (2)	3.98e-07 (17)	0.092733	list csv
* 21	13.00	●	1.36e-01 (7)	3.71e-07 (19)	0.009184	list csv
* 23	13.50	●	2.52e-02 (4)	3.49e-07 (23)	0.009009	list csv
* 25	13.50	●	1.18e-01 (6)	3.66e-07 (21)	0.007742	list csv
* 14	14.00	●	4.63e-01 (14)	4.14e-07 (14)	0.012654	list csv
* 10	19.50	●	7.82e-01 (37)	8.74e-06 (2)	0.021017	list csv
* 28	20.00	●	3.83e-01 (12)	3.05e-07 (28)	0.005736	list csv

Figure a. Full length barcode (660bp)

ABGD

Partition 1 : found 181 groups (prior maximal distance P= 0.001000)
Partition 2 : found 21 groups (prior maximal distance P= 0.001668)
Partition 3 : found 17 groups (prior maximal distance P= 0.002783)
Partition 4 : found 15 groups (prior maximal distance P= 0.004642)
Partition 5 : found 15 groups (prior maximal distance P= 0.007743)
Partition 6 : found 11 groups (prior maximal distance P= 0.012915)
Partition 7 : found 9 groups (prior maximal distance P= 0.021544)
Partition 8 : found 9 groups (prior maximal distance P= 0.035938)
Partition 9 : found 9 groups (prior maximal distance P= 0.059948)

Figure b. Full length (660bp)

Nb of species	asap-score	P-val (rank)	W (rank)	Treshold dist.	Text
9	2.00	1.00e-05 (3)	1.48e-04 (1)	0.055672	list csv
9	2.50	1.00e-05 (2)	3.67e-05 (3)	0.088742	list csv
* 11	6.00	2.89e-02 (6)	1.51e-06 (6)	0.016321	list csv
* 12	7.00	1.87e-01 (9)	1.51e-06 (5)	0.014649	list csv
* 10	7.50	6.23e-01 (13)	8.16e-05 (2)	0.021337	list csv
4	9.00	1.00e-05 (1)	3.91e-07 (17)	0.145450	list csv
* 22	11.00	3.51e-01 (12)	5.28e-07 (10)	0.008118	list csv
4	11.50	8.18e-01 (19)	1.24e-05 (4)	0.209741	list csv
* 17	12.00	1.86e-01 (8)	4.48e-07 (16)	0.009754	list csv
9	12.50	6.25e-01 (14)	5.02e-07 (11)	0.105754	list csv

Figure c. 307bp mini-barcode

Partition 1 : found 114 groups (prior maximal distance P= 0.001000)
Partition 2 : found 114 groups (prior maximal distance P= 0.001668)
Partition 3 : found 114 groups (prior maximal distance P= 0.002783)
Partition 4 : found 13 groups (prior maximal distance P= 0.004642)
Partition 5 : found 10 groups (prior maximal distance P= 0.007743)
Partition 6 : found 10 groups (prior maximal distance P= 0.012915)
Partition 7 : found 9 groups (prior maximal distance P= 0.021544)
Partition 8 : found 9 groups (prior maximal distance P= 0.035938)
Partition 9 : found 9 groups (prior maximal distance P= 0.059948)
Partition 10 : found 4 groups (prior maximal distance P= 0.100000)

Figure d. 307pb mini-barcode

Nb of species	asap-score	P-val (rank)	W (rank)	Treshold dist.	Text
* 11	2.00	1.00e-05 (3)	2.45e-04 (1)	0.047759	list csv
11	2.00	1.00e-05 (2)	2.51e-05 (2)	0.075082	list csv
63	9.00	1.20e-01 (7)	8.35e-07 (11)	0.003185	list csv
65	11.00	2.83e-01 (12)	8.43e-07 (10)	0.003178	list csv
71	11.00	2.91e-01 (13)	8.43e-07 (9)	0.003173	list csv
* 13	11.50	5.01e-01 (17)	1.00e-06 (6)	0.013566	list csv
* 12	12.00	5.61e-01 (21)	8.17e-06 (3)	0.020416	list csv
* 24	13.00	2.83e-01 (11)	7.76e-07 (15)	0.007345	list csv
84	13.50	5.11e-01 (19)	8.44e-07 (8)	0.003165	list csv
* 12	14.50	1.30e-04 (4)	6.29e-07 (25)	0.015151	list csv

Figure e. 313bp mini-barcode

Partition 1 : found 85 groups (prior maximal distance P= 0.001000)
Partition 2 : found 85 groups (prior maximal distance P= 0.001668)
Partition 3 : found 85 groups (prior maximal distance P= 0.002783)
Partition 4 : found 14 groups (prior maximal distance P= 0.004642)
Partition 5 : found 11 groups (prior maximal distance P= 0.007743)
Partition 6 : found 11 groups (prior maximal distance P= 0.012915)
Partition 7 : found 9 groups (prior maximal distance P= 0.021544)
Partition 8 : found 9 groups (prior maximal distance P= 0.035938)
Partition 9 : found 9 groups (prior maximal distance P= 0.059948)
Partition 10 : found 9 groups (prior maximal distance P= 0.100000)

Figure f. 313pb mini-barcode

Nb of species	asap-score	P-val (rank)	W (rank)	Treshold dist.	Text
* 10	2.00	2.00e-05 (3)	1.23e-04 (1)	0.047586	list csv
10	2.50	1.00e-05 (2)	2.49e-05 (3)	0.072379	list csv
* 11	5.00	1.29e-03 (5)	5.67e-06 (5)	0.019266	list csv
* 11	5.50	4.87e-01 (9)	7.09e-05 (2)	0.026442	list csv
* 16	6.50	5.51e-03 (6)	6.91e-07 (7)	0.012042	list csv
4	8.50	8.80e-01 (13)	1.56e-05 (4)	0.179567	list csv
4	9.00	1.00e-05 (1)	3.29e-07 (17)	0.134987	list csv
36	9.50	2.06e-02 (7)	4.86e-07 (12)	0.004819	list csv
* 12	10.00	6.01e-01 (10)	5.37e-07 (10)	0.014458	list csv
* 17	10.50	4.65e-01 (8)	4.38e-07 (13)	0.009639	list csv

Figure g. 208pb mini-barcode

Partition 1 : found 79 groups (prior maximal distance P= 0.001000)
Partition 2 : found 79 groups (prior maximal distance P= 0.001668)
Partition 3 : found 79 groups (prior maximal distance P= 0.002783)
Partition 4 : found 79 groups (prior maximal distance P= 0.004642)
Partition 5 : found 18 groups (prior maximal distance P= 0.007743)
Partition 6 : found 12 groups (prior maximal distance P= 0.012915)
Partition 7 : found 10 groups (prior maximal distance P= 0.021544)
Partition 8 : found 9 groups (prior maximal distance P= 0.035938)
Partition 9 : found 9 groups (prior maximal distance P= 0.059948)
Partition 10 : found 9 groups (prior maximal distance P= 0.100000)

Figure h. 208pb mini-barcode

Nb of species	asap-score	P-val (rank)	W (rank)	Treshold dist.	Text
9	2.00	1.00e-05 (3)	1.37e-04 (1)	0.054174	list csv
9	2.00	1.00e-05 (2)	1.26e-04 (2)	0.085408	list csv
* 11	8.50	2.50e-01 (12)	6.77e-07 (5)	0.013531	list csv
* 22	9.50	7.98e-02 (6)	4.86e-07 (13)	0.007642	list csv
* 12	13.50	2.97e-02 (4)	4.04e-07 (23)	0.011819	list csv
* 10	16.50	6.43e-01 (30)	1.54e-05 (3)	0.019162	list csv
123	17.00	4.03e-02 (5)	3.24e-07 (29)	0.002393	list csv
* 21	17.50	4.97e-01 (21)	4.85e-07 (14)	0.008217	list csv
* 27	18.00	3.99e-01 (18)	4.62e-07 (18)	0.006778	list csv
4	18.50	1.00e-05 (1)	2.98e-07 (36)	0.141069	list csv

Figure i. 361pb mini-barcode

Partition 1 : found 124 groups (prior maximal distance P= 0.001000)
Partition 2 : found 124 groups (prior maximal distance P= 0.001668)
Partition 3 : found 19 groups (prior maximal distance P= 0.002783)
Partition 4 : found 14 groups (prior maximal distance P= 0.004642)
Partition 5 : found 10 groups (prior maximal distance P= 0.007743)
Partition 6 : found 10 groups (prior maximal distance P= 0.012915)
Partition 7 : found 9 groups (prior maximal distance P= 0.021544)
Partition 8 : found 9 groups (prior maximal distance P= 0.035938)
Partition 9 : found 9 groups (prior maximal distance P= 0.059948)
Partition 10 : found 4 groups (prior maximal distance P= 0.100000)

Figure j. 361pb mini-barcode

Figure S1. A list of the "best" partitions (10 by default) generated by ABGD and ASAP. List of partition (A), with the number of species they correspond to, ranked by their ASAP score (B), provided in the output together with their p-value (C), their gap-width score W (D), their threshold distance (E). The darker the node color of the partition is, the best is the asap score. The smallest p-value has rank one, the largest gap has rank one (given in parenthesis) (Puillandre *et al*., 2021). ABGD reported a list of best partitions with the number of clusters in each partition according to a fixed range of prior maximum divergence of intraspecific diversity (P) ((Puillandre *et al*., 2021).

Table S1. MOTUs and sequences distribution among ASAP, ABGD and NJ methods

Method	MOTUs	Number of sequences
Dataset (BINs)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86
	MOTU4 : *Microgaster*	105
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	66 5
	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	302
ASAP (Full barcode)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86
	MOTU4 : *Microgaster*	105
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	66 5
	MOTU7: *Diolcogaste*	7

	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	302
ABGD (Full barcode)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86
	MOTU4 : *Microgaster*	105
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	66 5
	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	302
NJ (Full barcode)	MOTU1: *Glyptapanteles* MOTU1: Unidentified	40 1
	MOTU2: *Apanteles* MOTU2: Unidentified	185 2
	MOTU3: *Lissonota*	86
	MOTU4: *Microgaster*	104
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	106 5

	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	302
ASAP (310bp)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86
	MOTU4: Microgaster	105
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	66 5
	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	302
ABGD (310bp)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86
	MOTU4: *Microgaster*	105
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	106 5

	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	302
NJ (310bp)	MOTU1: *Glyptapanteles* MOTU1: Unidentified	40 1
	MOTU2: *Apanteles* MOTU2: Unidentified	185 2
	MOTU3: *Lissonota*	86
	MOTU4: *Microgaster*	104
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	106 5
	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	303
ASAP (313bp)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86
	MOTU4: *Microgaster*	105
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta*	106

	MOTU6: Unidentified	5
	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	300
	MOTU 10: *Dolichogenidea*	1
	MOTU11: *Dolichogenidea*	1
ABGD (313bp)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86
	MOTU4: *Microgaster*	105
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	106 5
	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	302
NJ (313bp)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86
	MOTU4: *Microgaster*	105

	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	106 5
	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	302
ASAP (208bp)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86
	MOTU4: *Microgaster*	105
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	66 5
	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	301
	MOTU 10: *Dolichogenidea*	1
ABGD (208bp)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86

	MOTU4: *Microgaster*	105
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	106 5
	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	302
NJ (208bp)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86
	MOTU4: *Microgaster*	105
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	106 5
	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	302
ASAP (361bp)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86

	MOTU4: *Microgaster*	105
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	66 5
	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	302
ABGD (208bp)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86
	MOTU4: *Microgaster*	105
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	106 5
	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	302
NJ (208bp)	MOTU1: *Glyptapanteles*	40
	MOTU2: *Apanteles* MOTU2: Unidentified	147 1
	MOTU3: *Lissonota*	86

	MOTU4: *Microgaster*	104
	MOTU5: *Cotesia*	106
	MOTU6: *Glypta* MOTU6: Unidentified	66 5
	MOTU7: *Diolcogaste*	7
	MOTU8: *Myrmica*	174
	MOTU 9: *Dolichogenidea*	302

Table S2. Results of molecular vs morphological identification of Hymenoptera

Sequencing sample ID	Barcode length	Molecular identification		Morphological identification	Mislabeled
		GenBank	B0LD		
MorphOTU1	Mini-barcode	Chalcidoidea Eulophidae (96,02%)	Chalcidoidea (98.97%)	Chalcidoidea	Correctly labelled
MorphOTU2	Mini-barcode	Diapriidae (96.3%)	Diapriidae (98.23%)	Diapriidae	Correctly labelled
MorphOTU2	Full barcode	Diapriidae (98.17%)	Diapriidae (98.40%)	Diapriidae	Correctly labelled
MorphOTU6	Mini-barcode	Diapriidae (98,34%)	Diapriidae (100%)	Proctotrupoidea	+
MorphOTU6	Full barcode	Diapriidae (99.93%)	Diapriidae (99.13%)	Proctotrupoidea	+
MorphOTU7	Mini-barcode	Pteromalidae (96.86%)	Pteromalidae (98.47%)	Pteromalidae	Correctly labelled
MorphOTU7	Full barcode	Pteromalidae (97.93%)	Low confidence	Pteromalidae	Correctly labelled
MorphOTU17	Mini-barcode	Ichneumonidae Cryptinae (98.53%)	Ichneumonidae Cryptinae (98.53%)	Ichneumonidae Cryptinae	Correctly labelled

MorphOTU14	Mini-barcode	Diapriidae (97.62%)	Diapriidae (97.87%)	Diapriidae	Correctly labelled
MorphOTU14	Full barcode	Diapriidae (98.07%)	Low confidence	Diapriidae	Correctly labelled
MorphOTU12	Mini-barcode	Platygastridae (97.31%)	Low confidence	Platygastridae	Correctly labelled
MorphOTU12	Full barcode	Platygastridae (98.99%)	Platygastridae (99.01%)	Platygastridae	Correctly labelled
MorphOTU15	Full barcode	Ichneumonidae *Pimpla* (99.24%)	Ichneumonidae *Pimpla* (99.24%)	Ichneumonidae	Correctly labelled
MorphOTU16	Full barcode	Ichneumonidae Orthocentrinae (99.08%)	Ichneumonidae Orthocentrinae *Orthocentrus* (99.53%)	Ichneumonidae	Correctly labelled
MorphOTU22	Full barcode	Diapriidae (97.01%)	Diapriidae (96.09%)	Diapriidae	Correctly labelled
MorphOTU22	Mini-barcode	Diapriidae (95.64%)	Diapriidae (98.17%)	Diapriidae	Correctly labelled
MorphOTU18	Full barcode	Ichneumonoidea Tersilochinae (97.17%)	Ichneumonoidea (97.96%)	Ichneumonoidea	Correctly labelled
MorphOTU29	Full barcode	Chalcidoidae	Ceraphronidae (96.61%)	Ceraphronidae	Correctly labelled/+

		Encyrtidae (95.17)			
MorphOTU10	Mini-barcode	Diapriidae (95.94%)	Diapriidae (96.54%)	Diapriidae	Correctly labelled
MorphOTU9	Mini-barcode	Ceraphronidae (96.18%)	Low confidence	Ceraphronidae	Correctly labelled
MorphOTU28	Mini-barcode	Platygastroidae Scelionidae (98.46%)	Platygastroidae Scelionidae (96.58%)	Platygastroidea Scelionidae	Correctly labelled
MorphOTU26	Mini-barcode	Diapriidae (95.94%)	Low confidence	Diapriidae	Correctly labelled
MorphOTU2	Mini-barcode	Diapriidae (95.19%)	Low confidence	Diapriidae	Correctly labelled
MorphOTU31	Mini-barcode	Platygastroidea Scelionidae (98.07%)	Low confidence	Platygastroidea Scelionidae	Correctly labelled
MorphOTU27	Mini-barcode	Diapriidae (95.02%)	Diapriidae (96,03%)	Chalcidoidae	+
MorphOTU26	Mini-barcode	Diapriidae (95.48%)	Low confidence	Diapriidae	Correctly labelled
MorphOTU19	Mini-barcode	Diapriidae (96.89%)	Low confidence	Diapriidae	Correctly labelled

MorphOTU21	Mini-barcode	Diapriidae (97.02%)	Low confidence	Diapriidae	Correctly labelled
MorphOTU9	Mini-barcode	Ceraphonidae (95.34%)	Ceraphronidae (95.52%)	Ceraphronidae	Correctly labelled
MorphOTU25	Mini-barcode	Platygastroidea Scelionidae (96.63%)	Low confidence	Platygastridae	Correctly labelled

Chapter 3

Multiplex High throughput DNA barcoding to assess Hymenoptera diversity in the Laurentian Forest –What's the congruence between morphological and molecular identification approaches?

Multiplex High throughput DNA barcoding to assess Hymenoptera diversity in the Laurentian Forest –What's the congruence between morphological and molecular identification approaches?

Malek. Kalboussi[1], Colin. Favret[1]

[1] University of Montreal, Plant Biology Research Institute (IRBV), Biology Centre, 4101rue Sherbrooke Est, Montréal, Québec, H1X 2B2, Canada.

Authors contribution

Malek. Kalboussi[1,] Conception, reaserch work and writing

Colin. Favret[1,] Conception, editing, and revision

Corresponding author. Malek kalboussi. Email: malek.kalboussi@umontreal.ca

In preparation to be submitted

Abstract

The two major challenges to identifying insects are their high diversity and the large number of specimens sampled. Taking into account the continuing loss of taxonomic expertise also known as the“ Taxonomic impediment”, species identifications are difficult to achieve, especially for the smallest, most diverse insects. Among these smallest of insects, Hymenoptera are poorly known and their diversity is underestimated in Canada. In order to address the challenges to insect biodiversity research, instead of trying to identify all specimens to species, Morphological Operational Taxonomic Units (MorphOTUs), also known as morphospecies, are often used. Alternately, molecular OTUs based on short DNA sequences (often called DNA barcodes) can be used. It remains an open question as to which method of operational taxonomic classification performs better for evaluating the diversity of Hymenoptera. We tested a MorphOTU methodology against a MOTU methodology on 517 specimens of suction-trapped Hymenoptera from the Laurentian forest of eastern Canada. DNA mini-barcodes, 313bp long, were acquired by multiplex Illumina sequencing. These were subjected to several species delimitation algorithms and the resulting clusters compared with those created by student parataxonomists who had sorted the same specimens into MorphOTUs.

128 MorphOTUs were classified. Because we were unable to obtain some DNA sequences, we only tested 103. The four molecular species delimitation methods we tested (distance-based, Automatic Barcode Gap Discovery (ABGD) and Assemble Species by Automatic Partitioning (ASAP), and tree based Generalized Mixed Yule Coalescent (GMYC) and the Bayesian Poisson Tree Processes (bPTP)), gave divergent results from the morphological classification. ASAP recovered MOTUs most closely aligned with the MorphOTUs, followed by ABGD, but tree-based methods tended to increase the number of MOTUs by over-splitting relative to distance-based methods. The overall difference between MorphOTUs and molecular delimitation varies among Hymenoptera super-families. It was most conserved within Chalcidoidea and Ceraphronoidea.

Regarding the obvious incongruence between morphological and molecular delimitation, we were not able to judge the effectiveness of each method except based on time and cost, which were lower with molecular classification. It will be interesting to compare the fit-for-purpose performance of the two kinds of datasets in an actual actual biodiversity assessment.

Keywords: morphology, species delimitation, COI.

1. Introduction

Insects are the most diverse class in the animal kingdom (Chapman, 2009; Yeo *et al*., 2019; Bennett *et al*., 2019a) with over 1 million identified species (Mora *et al*., 2011; Hofreiter *et al*., 2015; Stork.,2018; Patzold *et al*., 2020) and millions more to be discovered (Gilbert *et al*., 2007; Zhang, 2011; Stork, 2018; Yeo *et al*., 2020). Trap catches reveal the enormous number of insects sampled for many groups (Favret and Voegtlin, 2004; Sheikh *et al*., 2016; Favret *et al*., 2019).

Sorting of such samples has traditionally been done based on morphology, for insect ecology, evolution, conservation biology, species identification, biodiversity assessment and biomonitoring research (Ratnasingham and Hebert 2007; Floyd *et al*., 2010; Yu *et al*., 2012; Hebert *et al*., 2016; Blagoev *et al*., 2015; Ashfaq and Hubert, 2016; Bennett *et al*., 2019a; Srivathsan *et al*., 2021). However, it is not simple to recognize, name, and identify species: it takes time and taxonomic expertise (Krell; 2004; Favret *et al*. 2019; Yeo *et al*., 2020), making many diverse insect groups poorly known by traditional taxonomy (Abadie *et al*., 2008; Stork *et al*., 2018; Bar-On *et al*., 2018).

To overcome this“ Taxonomic impediment” (Vinarski, 2020) in ecological biodiversity research, parataxonomists often sort samples to recognized taxonomic units (RTUs) (Oliver and Beattie 1993), also known as morphospecies (Derraik *et al*., 2002), morphotypes (Abadie; 2008), parataxonomic units (PUs) (Krell, 2004) or, as proposed here, morphological operational taxonomic units (MorphOTUs). MorphOTUs are putatively different taxa separated based on apparent morphological characteristics (Derraik *et al*., 2001; Hebert *et al*., 2016; Meier *et al*.,2016; Wang *et al*., 2018). They have been widely regarded as a sufficient and trustworthy approach to accelerate biodiversity assessment (Derraik *et al*., 2002; Saunders, 2018; Favret *et al*. 2019), saving time and cost. However, sorting can be imprecise and potentially erroneous, especially with small, cryptic insect groups (Krell, 2004; Favret *et al*., 2019; Yeo *et al*., 2020).

With approximately 160,000 species in 94 extant families, Hymenoptera (Stahlhut *et al*., 2013; Barroso *et al*., 2022) is one of the largest insect orders (Forbes *et al*., 2018). Despite their fundamental role as parasitoids, predators, pollinators and bio-indicators in all terrestrial ecosystems (Sheffield *et al*., 2009; Peters *et al*., 2017), many families of Hymenoptera are inadequately known, both at the global level (Forbes *et al*., 2018; Belokobyl’skij *et al*., 2019) and in Canada (Hebert *et al*., 2016; Peters *et al*., 2017). Several publications have remarked on the great morphological homogeneity within many Hymenoptera groups (Rasnitsyn *et al*.,

1988; LaSalle, 1993; Austin and Dowton, 2000; Bennett *et al*., 2019b; Parslow *et al*., 2021; Barroso *et al*., 2022), which makes species identification challenging.

One solution is to use molecular data to define Molecular Operational Taxonomic Units (MOTUs) through a DNA barcoding approach (Wong *et al*., 2014; Wang *et al*., 2018; Parslow *et al*., 2021; Barroso *et al*., 2022). Research on multiple Hymenoptera families has employed COI barcodes for species delimitation, including hyper diverse groups like Braconidae (Zaldivar-Riverón *et al*., 2010; Fagan-Jeffries *et al*., 2018), Formicidae (Oberprieler *et al*., 2018) and Vespidae (Barroso *et al*., 2022).

In large-scale investigations using DNA barcoding to assist morphological identification, researchers often start their workflow with morphology-based pre-sorting to different levels (family, genus, species) according to their budget and available taxonomic expertise (Renaud *et al*., 2012; Hebert *et al*., 2016; Wang *et al*., 2018). They define MorphOTUs and barcode a few specimens of each. Even the most ambitious barcoding study to date (Hebert *et al*., 2016) used presorting for (>1 million barcodes) to assess the diversity of 27 orders of Canadian insects with several million sampled specimens. The Canadian Centre for DNA Barcoding in Guelph, Canada, is the best-known and well-funded institution in this field (Wang *et al*., 2018; Srivathsan *et al*., 2021). It generated more than 85% of all arthropod barcodes in the Barcode of Life Datasystems (BOLD) (Srivathsan *et al*., 2021; BOLD Systems, 2022).
Unfortunately, this hybrid approach of barcoding only exemplars of morphospecies (Srivathsan *et al*., 2021) is not ideal for biodiversity assessment. Firstly, the validity and accuracy of morphospecies are inconsistent and unpredictable when presorting is done by parataxonomists (Krell *et al*., 2004; Chang *et al*., 2012), which can carry major repercussions (mainly cryptic species, incorrect morphospecies sorting, etc) (Wang *et al*., 2016; Favret *et al*., 2019). Secondly, barcoding fewer samples is not credible to judge the level of mismatch between molecular and morphological identification, since both forms of data are not used to study all specimens.

Alternately, researchers can reverse the workflow by barcoding all the specimens first and then analyzing their morphological congruence, following a morphological identification of representative morphospecies (Wang *et al*., 2018; Srivathsan *et al*., 2019; Srivathsan *et al*., 2021). This process breaks the dependence between molecular and morphological data (Morinière *et al*., 2016; Wang *et al*., 2018), especially for bulk samples, when thousands or even millions of specimens are processed in a metabarcoding approach (Zhang *et al*., 2018). Yet, these methods are not without flaws. Firstly, voucher specimens can be either entirely or

partially destroyed during DNA extraction. Secondly, during the bioinformatics analysis, most barcodes cannot be determined to species using online databases (e.g., BOLD, GenBank), especially for under-described groups such as Hymenoptera (Bennett *et al*., 2019a). As a result, determining whether similar reads in the sample are from sequence variation (haplotypes) in a single species or from closely-related but distinct species is difficult, and is especially problematic in the case of rare species (Batovska *et al*., 2021). It is unclear how much this meta-barcoding data can reflect the precise abundance of samples for correct biodiversity assessment, since most metabarcoding pipelines are not robust enough to return reliable results (Shokralla *et al*., 2015; Meier *et al*., 2016). Lastly, we still have the same problem of insufficient morphological data as much as molecular data.

To our knowledge there has been no study that compared the effectiveness of molecular vs morphological operational taxonomic units to evaluate the biodiversity of Hymenoptera in Canada. Taking into account the current number of undescribed species, in the world generally and in Canada, there is a need to explore the efficiency of the MOTUs to discriminate between species in order to evaluate the biodiversity of Hymenoptera.

The objective of this study was to explore to utility of high-throughput DNA barcoding on hundreds of Hymenoptera as compared to morphology-based classification. We developed and optimized a High throughput DNA mini-barcoding approach, including its several laboratory steps; DNA extraction, PCR amplification and high-throughput sequencing (HTS). Subsequently, we compared commonly used molecular delimitation approaches to morphological assignation, studying the congruence between molecular and morphological operational taxonomic units.

2. Materials and methods

2.1.Hymenoptera sampling

Hymenoptera samples were collected as part of an insect survey based on suction trapping (Favret *et al*., 2019) that was conducted over six consecutive months, May-October 2021, and included four different sites at the University of Montreal's Laurentian Biology Research Station (St-Hippolyte, Quebec; 45.99, –74.01) (Savage, 2001) (Figure S2). Each of the four sites has a high resolution panoramic image provided by Favret *et al*. (2019).

2.2.Post sampling process and samples used

The suction trap was checked every week when the collecting bottle with the material was replaced. Each suction trap sample was assigned a unique registration number to make specimen tracking easier. Subsequently, the material was collected and transferred to an Eppendorf tube, stored at 70% ethanol, and pre-sorted to insect order or family based on morphology (Figure S3). We here used 517 Hymenoptera specimens from a single weekly sample (1-8 July 2021) as recommended for rapid insect biodiversity assessment (Ritter *et al*., 2019).

2.3.Morphological sorting of Hymenoptera and MorphOTU assignation

Target taxa belonging to Hymenoptera were sorted to family level using relevant taxonomic keys (Yoshimoto, 1984; Goulet and Hubert, 1993; Belokobyl'skij and Lelej, 2019). Specimens were further sorted to MorphOTUs, so that the different catches would be comparable (Figure S3). A variety of subjectively chosen characteristics were used to sort MorphOTUs: color, body, leg and antenna length, and wing venation complexity were of particular focus in sorting out with different MorphOTUs of Hymenoptera. Finally, specimens were examined under a stereomicroscope with magnification for MorphOTU verification, photographed. Photographs and an interactive identification key developed in-house were consulted to ensure standard specimen interpretation. A team of six undergraduate interns worked together and every specimens was cross-validated between them to ensure consistent classification.

2.4.Multiplex DNA barcode sequencing

2.4.1. DNA extraction

We used QuickExtract (Lucigen, NYC, USA) to extract DNA individually from 517 specimens in six 96-well plates. The specimens were immersed in ultrapure water for 20min before being air-dried on sterile filter paper prior to DNA extraction using the previously-described protocol (Chapter 2). Whole specimens were immersed in a volume ranging from 20 to 50µL of QuickExtract solution (depending on the specimen size). DNA was obtained within 17min in a

thermocycler via two heating steps; 65°C for 15min followed by 98°C for 2min. The DNA concentration and absorbance ratio (A260/A280) of specimens were measured individually by spectrophotometry (NanoPhotometerP330, Canada).

2.4.2. PCR amplification of mini-barcodes and library preparation

A 313bp fragment of COI of the 517 Hymenoptera specimens was amplified using labeled forward and reverse primers; m1COlintF 5'- GGWACWGGWTGAACWGTWTAYCCYCC-3' (Leray *et al*., 2013) and modified jgHCO2198 5'-TAAACYTCAGGRTGCCCRAARAAYCA-3' (Meier *et al*., 2016). We used the Meier *et al.* (2016) labels generated by the online freeware "Barcode Generator"; each label was 9bp long, differing from other labels by ≥4bp (Table S3). Touchdown (TD) PCR reactions were performed to increase specificity of PCR reactions (Green and Sambrook, 2018) and avoid primer mismatch with the different indexed primers (Korbie and Mattick, 2008). Three different touchdown reactions were tested with three different temperatures: 60-50°C, 62-55°C, 65-55°C. TD 60-50°C provided a 100% amplification rate and high intensity electrophoresis bands. We started with a higher annealing temperature (60°C) which then is gradually decreased until it gets closer to the perfect annealing temperature of primer pair m1COlintF—Modified jgHCO2198, that is, 50°C. PCR amplification conditions were 50s initial denaturation at 94°C followed by 35 cycles of denaturation at 94°C for 45sec, annealing temperature ranged between 60°C and 50°C for 1min, extension at 72°C for 1min, followed by final extension of 72°C at 5min. We used the following reaction mix: 10µL of Phire Green Hot Start II PCR Master Mix (Thermo Fisher Scientific, Waltham, MA), 2µL each of 10µM tagged primers, 2µL of template and DNase-free sterile water up to 20µL.

We assigned the primer tag combinations as follows. For each plate with 96 PCR reactions, we added the same Forward (F) primer for one vertical row; 12 different F-primers for 12 rows (Figure S5). The same tagged Reverse (R) primers were added in the same horizontal row; eight different tagged R primers for eight rows. We than added the master mix, water and the DNA template. All 96 samples in a plate have a unique combination of tagged primers, with a negative control to detect contamination. The six plates share the same 12 unique tagged F primers and eight unique tagged R primers. All amplifications were carried out using a Thermocycler Eppendorf Matercycler ProS (Eppendorf, Canada). The 73 to 96 amplicons of each of the six plates were pooled (Figure S5) and the six resulting samples submitted to Genome Quebec for library preparation, cleaning, normalization, and sequencing on Illumina Miseq (2*250PE) (Illumina, Quebec, CA, USA). Our library occupied 1/6 of the lane. A total

of 519 samples (517 Hymenoptera amplicons and two negative controls with only PCR reagents) were sequenced. Each plate had its unique Illumina sequencing adapters (Pease and Sooknanan, 2012) that were used to demultiplex the reads corresponding to each pool obtained during sequencing.

Mean consumable cost and time of High throughput DNA barcoding, including each different step, were evaluated to be compared to the MorphOTUs classification.

2.5.Bioinformatic pipeline

FASTQC software from the free online Galaxy/ Pasteur platform (https://galaxyproject.org/use/galaxy-pasteur/) was used to examine read quality. Sequences were edited and reassembled using Geneious Prime® 2022.1.3 software (Kearse *et al.*, 2012). A pipeline was created for the following steps. This software required a fastq file of both ends of target mini-barcodes, generated by Ilumina Miseq sequencer (2*250 PE). Reads R1 and R2 were merged using BBMerge from the BBtools suite. Geneious was used to demultiplex sequences, allowing PE reads to be attributed to their appropriate specimen of origin. For demultiplexing, no mismatch was allowed for the tag region while two mismatches were allowed for the primer sequence. Trimming of low quality ends of sequences, removing of Illumina adaptors and primers, and filtering out demultiplexed reads by length was performed using BBDuk trimmer from Geneious. A set of parameters was fixed based on research recommendations (Bokulich *et al*., 2013; Staats *et al*., 2016); a minimum threshold in quality filtering was set to 20 phred score (Staats *et al*., 2016), short reads of minimum length of 10 were discarded, error probability limit was set to 0.05. Barcodes were assembled *de novo* and to a reference sequence belonging to a platygastroid. Sequences were examined to check the coding frame for possible stop codons. Sequences were aligned with the ClustalW tool from MEGA X (Molecular Evolutionary Genetics Analysis Version X) software (Edgar, 2004) with default option settings. All alignments were visually verified and rectified in Geneious.

2.6.Molecular OTU identification

Mini-barcodes were matched to reference sequences via a BLAST check in GenBank (NCBI) and the online Barcode of Life Database (BOLD-Ratnasingham and Hebert, 2007). Barcodes that showed an identification rate > 98% match to the closest library sequence were assigned a species-level identification (Hebert *et al*., 2003), We choose a threshold >97% similarity to confirm genus-level >96% family-level, and 85% for order-level based on literature (Meier *et al*., 2016; Wang *et al*., 2018; Piper *et al*., 2019).

2.7.Molecular OTU estimation

2.7.1. Distance-based delimitation methods

The automatic barcode gap discovery method (ABGD) (Puillandre *et al*., 2012) and Assemble Species by Automatic Partitioning ASAP (Puillandre *et al*., 2021) were used to test two distinct distance-based species delimitation methodologies through the online platforms available at https://bioinfo.mnhn.fr/abi/public/abgd and https://bioinfo.mnhn.fr/abi/public/asap. The algorithms require an alignment as an input to produce a distance matrix employing the Kimura Two-Parameter (K2P) model (Kimura 1980). ABGD and ASAP use the distribution of pairwise distances to find sequence clusters. These time-saving methods divide the data recursively and compare the differences across sequences to find a "barcode gap" that may represent species boundaries (Puillandre *et al*., 2012, 2021). We ran ABGD with the following settings: Pmin = 0.005, Pmax = 0.1, and 20 steps, with gap width=1. The MOTU assignments were recorded over 20 recursions.

2.7.2. Tree-based delimitation methods

2.7.2.1. Phylogenetic tree construction

Fifteen insuitable Hymenoptera haplotypes were discarded from a set of 458 mini-barcodes for the phylogenetic analysis. To obtain the best phylogenetic tree model, we first inferred phylogenies with maximum likelihood (ML) using the open webserver of phyML3.0 software (Guidon *et al*., 2010), available at: http://www.atgc-montpellier.fr/phyml/results.php?path=20220520-193437_Xm69/sms.csv. The unrooted ML tree search was performed following a best Model Finder of sequence evolution under the Bayesian information criterion (BIC) (Lefort *et al*., 2017). BEAUTi and BEAST 1.10.4 were used to estimate ultrametric trees required for tree-based species delimitation methods (Drummond *et al*., 2006), employing the model suggested by PyML3.0. We set the model of strict molecular clock, as it is recommended in (Michonneau, 2016) to calculate trees with empirical frequency-based priors starting with a random tree (Drummond *et al*., 2006). We calculate Yule tree prior model for species level data: this assumes a constant speciation rate to explain all branches in the tree (Willis, 1925). The default value of 10,000,000 was used for the number of Monte Carlo Markov Chains (MCMC) generations and every 10,000 trees were sampled. Tree Annotator v1.10.4 was used to produce the maximum credibility tree. The software FIGTREE v.1.4.2 (http://tree.bio.ed.ac.uk/software/fgtree) was used to visualize the phylogenetic trees.

2.7.2.2. Delimitation Software

We used two tree-based species delimitation methods. The single Threshold of the Generalized Mixed Yule Coalescent (sGMYC) analysis (Pons *et al*., 2006), and Bayesian Poisson Tree Processes (bPTP) (Zhang *et al*., 2013) to analyze the sequence data.

The generalized mixed Yule coalescent (GMYC) analysis is a coalescent-based phylogenetic method for defining species boundaries by establishing thresholds between coalescent and species-level processes (Fujisawa and Barraclough, 2013; Fontaneto *et al*., 2015). The only input required for this tree-based technique is the ultrametric phylogenetic tree with the maximum credibility generated by BEAST software. (Pons *et al*., 2006; Fujisawa and Barraclough., 2013). The SPLITS package on the R platform was used for the GMYC analysis, available at http://r208 forge.r-project.org/projects/splits/.

The Poisson Tree Processes (PTP) models speciation events in terms of the amount of substitutions in a given branch, implying that there would be more substitutions between species than within species (Zhang *et al*., 2013). For the Bayesian PTP approach, we employed phylogenetic trees constructed by BEAST. We used the python version of bPTP, deleting the outgroups and using default parameters, except for the number of Monte Carlo Markov Chains (MCMC) generations of 300,000.

2.8.Verification of overlap between morphological and molecular OTU

Validation of congruence between MOTU and MorphOTU was based on species delimitation and identification levels (family, genus, species). MorphOTUs initially sorted using relevant taxonomic keys were compared with matches from GenBank (NCBI) and BOLD following a BLAST search of mini-barcoded specimens.

For the MOTUs delimitation, the purpose was to check if the delimitation of molecular OTUs corresponded to morphological OTUs both within (in the same MOTU) and between MOTUs. The match ratio, modified from Ahrens *et al*., (2016), used by Wang *et al*. (2018), was considered to quantify the results: 2 Nmatch / (NMOTU + NMorphOTU), where Nmatch is the number of assigned clusters that are the same in both techniques of delimitation, and NMOTU and NMorphOTU are the total number of MOTUs and total number of morphological species units, respectively. Singleton sequences were excluded. Quantification of results was conducted on 64 MOTUs that had more than one sequence, as defined by molecular delimitation methods, with a total of 405 mini-barcodes.

3. Results

3.1. Morphological sorting of Hymenoptera and MorphOTUs assignation

In all, 128 MorphOTUs from 517 specimens were assigned. MorphOTUs were first identified to super-family and family using relevant taxonomic keys and reference collections (Yoshimoto, 1984; Goulet and Hubert, 1993; Belokobyl'skij and Lelej, 2019), followed by sorting to MorphOTU.

Chalcidoidea was the richest and most abundant Hymenoptera super-family in our samples (44 MorphOTUs): 13 MorphOTUs were identified as Mymaridae, two as Trichogrammatidae, and the rest kept at super-family level. In second place were Ichneumonoidea (16 Ichneumonidae and 13 Braconidae MorphOTUs) and Ceraphronoidea, the former being more diverse, the latter being more abundant (Table 3.1). We recorded also 13 Platygastroidea MorphOTUs, with four Scelionidae and nine Platygastridae, and 11 different Diaprioidea MorphOTUs (all Diapriidae). Hymenoptera of the super-families Cynipoidea (two Figitidae) and Vespoidea (three Pompilidae) were also recorded with lower abundance and richness. Eight MorphOTUs were left at the ordinal level.

Table 3.1. Numbers of MorphOTUs and specimens of Hymenoptera studied.

Hymenoptera super-family	MorphOTUs	Specimens
Chalcidoidea	44	147
Ichneumonoidea	29	78
Ceraphronoidea	18	112
Platygastroidea	13	72
Diaprioidea	11	64
Vespoidea	3	8
Cynipoidea	2	4
Unidentified	8	32
Total	128	517

3.2.High throughput DNA barcoding

3.2.1. DNA extraction

The DNA extracted from samples was generally high quality. The DNA yield of most specimens varied from 500 and 3506ng with an average of 1091,7±0.41ng. The mean absorbance ratio A260/A280 of the extracted DNA was 1.809±0.032 indicating a good quality for the most of specimens. Extractions from large specimens gave lower or higher absorbance ratios (<1.9 or >2) indicating a likely contamination with proteins (Chen *et al*., 2010) or RNA (Wang *et al*., 2019).

3.2.2. PCR amplification

The amplification success rate was 92%. An example of electrophoresis profile is presented in Figure 3.1. Low-intensity bands were obtained mainly with 7.4% of very small Chalcidoidea and 28% of Ceraphronoidea.

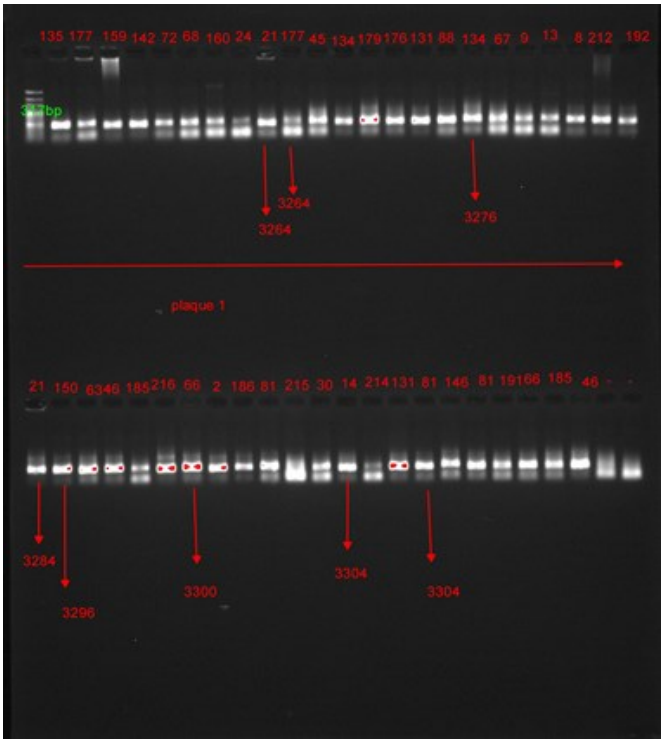


Figure 3.1. PCR amplification of mini-barcode COI from samples of Hymenoptera

3.2.3. Multiplex HTS

Barcoding was performed on a total of 517 Hymenoptera. In total 254,198 million reads were generated. Barcodes sequencing success rate was on average 93,3% per run. In all, 489 mini-barcodes were successfully sequenced with lower ambiguities (<10 bad quality nucleotides among the total length of mini-barcodes). Mean coverage was 83,88 X reads per specimen. We found a problem with the tag R1 “caacagtag”, mainly responsible for the loss of recorded

sequences during the demultiplexing step. After quality filtering and coding frame verification, a total of 458 mini-barcodes were considered.

DNA barcoding was 100% successful for Ichneumonoidea, Platygastroidea and Diaprioidea. We failed to obtain 50%, 23% and 16% of sequences respectively for Vespoidea, Ceraphronoidea and Chalcidoidea samples due to low read numbers (Table 3.2).

Table 3.2. DNA-barcode acquision rate according to Hymenoptera super-families

Hymenoptera super-families	Initial number of specimens	Number of sequences obtained by DNA barcoding	Percentage of DNA barcoding success
Chalcidoidea	147	142	84%
Ichneumonoidea	78	78	100%
Ceraphronoidea	112	87	77%
Platygastroidea	72	72	100%
Diaprioidea	64	64	100%
Vespoidea	8	4	50%
Cynipoidea	4	4	100%

3.3. Consumable costs: Barcoding vs. morphology-based identification

In all, 517 morphologically identified Hymenoptera were analyzed with DNA barcoding. This resulted in a general workload of up to 9h30min for a 96-well plate (the time for outsourced sequencing is not considered), with a material cost of 2.29CAD per specimen (Table 3.3). PCR amplification was the longest step. Ninety-six specimens were processed to PCR amplification in about 5 hours of work a day, including preparation of PCR reactions and verification of PCR amplification by gel electrophoresis. One person barcoded all 517 specimens in 35h spread over 6 days. About 110 specimens were re-amplified due to the failure of the first amplification. The

average cost of PCR consumables was 1.52CAD per specimen. HTS cost were about 0.44 per specimen. MorphOTUs assignation was done by a group of trainee students. Every Hymenoptera took an average of 10min for 5.83CAD per specimen, based on trainee`s salary.

Table 3.3. Comparison of time and cost effort for both molecular and morphological identification

Variable	**Multiplex DNA barcoding**				**Morphological classification to MorphOTUs**
	DNA extraction	**PCR amplification**	**HTS**	**Data analysis**	
Number of specimen simultaneously processed	96	96	96	96	1
Time	2h 30min	5h	+/- 45 days	2h	10min
Cost (CAD)	96* **0.35**	96***1.52**	96***0.44**	-	5.83
Total cost/specimen	**2.29**				**5.83**

3.4.Molecular OTUs identification

The 458 mini-barcodes were compared to the available sequences in GenBank and BOLD. Results yielded 458 identifications confidently assigned to a query sequence in GenBank and BOLD; these were identified to subfamily (376) with identification threshold >96% (for which reference sequences did not contain genus or species level identification in either database), to genus (75) with identification threshold >97%, and to species (7) with identification threshold >98%. Molecular identification in Genbank and BOLD databases revealed 133 Chalcidoidea specimens (Trichogrammatidae, Mymaridae, Pteromalidae, Eulophidae and Aphelinidae), 90 Ceraphronoidea (Ceraphronidae and Megaspilidae), 76 Ichneumonoidea (Ichneumonidae and Braconidae), 94 Platygastroidea (Platygastridae and Scelionidae), 51 Diaprioidea (Diapriidae),

4 Cynipoidea (Figitidae), and 4 Vespoidea (Pompilidae). Six sequences were identified as non-Hymenoptera in both databases. Molecular identification in BOLD was congruent with GenBank identification in 98,68% of sequences.

3.5. Molecular OTU delimitation

3.5.1. Distance-based delimitation methods

The intra-specific K2P distance for mini-barcodes COI was <5% for all the MOTUs under consideration. Inter-specific K2P was >5%.

ASAP delimitation resulted in 119 to 158 MOTUs under ten different partitions with ASAP scores of 9 to 31 (Figure S6). Only partitions with the lowest ASAP scores were considered as recommended by Puillandre *et al*. (2021). The lowest ASAP score recorded 137 clusters (Figure S6) at a prior distance of 0.05.

ABGD also recorded ten partitions. Only the initial and most stable partitions were considered (those with several genetic divergences (P) near the barcode gap), per Puillandre *et al*. (2021) (Figure S7). It outputted 174 clusters at four prior intraspecific divergences of (P)=0.006, 0.008, 0.093, and 0.01 and a relative gap width of (X)=0.039. The most stable partition has a tendency to overestimate MOTU numbers relative to the ASAP partition, at low prior intraspecific divergences of (P). Results were divergent from morphological OTUs estimation as well as ASAP partitions; consequently, we also considered the eighth partition on ABGD with 1,5% intraspecific threshold: it outputted 151 stable clusters at two prior intraspecific divergences of (P)=0.015 and 0.017, relative gap width (X)=0.039.

3.5.2. Tree-based delimitation methods

PhyML software suggested the GTR+R model as it was reported with the lowest BIC criterion (Table S2). The average length of 443 mini-barcodes was 332bp for sequenced Hymenoptera. In total 296 our of 380 sites (78%) were polymorphic.

The GMYC method recovered 222 MOTUs with the Yule prior model. It inferred 96 maximum likelihood (ML) clusters, that is species represented by at least two sequences, and 222 ML entities. This means that 123 MOTUs were represented by a singleton (Michonneau, 2017). bPTP took as input the same ultrametric tree generated by BEAST (Figure S8) and recognized 194 MOTUs for the Yule prior.

3.6.Overlap between morphological and molecular OTUs

Validation of congruence between putative species units based on DNA and morphology was carried out on two levels: species identification (via BLAST searches) and MOTUs delimitation.

3.6.1. MOTUs identification

The direct comparison of the previously generated morphological identification vs. BOLD and/ or GenBank revealed a 93.23% match at family and super-family levels. 18.5% of specimens were identified to higher level (genus, species) based on molecular data, permitting us to classify seven morphologically non-identified specimens.

3.6.2. MOTUs delimitation

Both distance-based (ABGD and ASAP) and tree-based (GMYC and bPTP) returned varying molecular operational taxonomic units (MOTUs), ranging from 137 to 222 (Table 3.4). Molecular methods tend to split MorphOTUs into multiple MOTUs. Over splitting referring to morphological delimitation, was most frequent with the tree-based analysis (Table S5, S6). Many of the MOTUs reflect singleton or otherwise rare species represented by few specimens (1 = 38 MOTUs, 2 = 19 MOTUs, 3= 8 MOTU and 4 = 7 MOTUs).

Table 3.4. Number of OTUs (MorphOTU or MOTU) in each Hymenoptera super-family under each delimitation method. Numbers in brackets refer to the number of specimens in the OTUs.

Délimitation Methods / **Hymenoptera super-families**	**MorphOTUs**	**ASAP**	**ABGD**	**GMYC**	**bPTP**
Chalcidoidea	34(132)	31	36	66	49
Ichenumonoidea	24 (74)	29	31	38	36
Diaprioidea	10(51)	22	23	28	28
Ceraphronoidea	14(89)	26	27	38	36
Platygastroidea	17(89)	24	28	46	40
Vespoidea	1(4)	2	2	3	2
Cynipoidea	2(4)	3	3	3	3
Total number of OTUs (Morphologial or Molecular)	102(443)	137	150	222	194

The total match ratio between morphological and molecular delimitation methods was lower than 0.35. There was a conflict between MorphOTU and MOTU estimation for 65% of cases. The congruence between morphological and molecular OTU was higher with ASAP and ABGD delimitation methods (with a match ratio of 0.34 and 0.28) than with tree-based methods (0.08 and 0.13 for GMYC and bPTP, respectively). Eleven MorphOTUs were fully congruent with MOTUs according to all four delimitation methods (Table S6) belonging to Chalcidoidea, Ichneumonoidea, Platygastroidea, Ceraphronoidea and Diaprioidea. ASAP, ABGD and bPTP have more overlap between MorphOTUs and MOTUs. These methods also support three to nine other MorphOTUs to be congruent with molecular identification (Table 3.5).

Table 3.5. Number of congruence MorphOTUs and MOTUs according to delimitation methods

Delimitation method	**ASAP**	**ABGD**	**GMYC**	**bPTP**
Number of congruent MorphOTUs	20	17	11	14

Based on ASAP, a total of 44 among 64 non-singleton MorphOTUs were without congruence to MOTUs (Figure 3.2). Eleven MorphOTUs were split into several MOTUs, whereas nine MorphOTUs cases were merged into the same MOTU with others. The majority of MorphOTUs (24) were considered mixed, meaning that some specimens from the same MorphOTU would be congruent, others are assigned to new MOTUs (Figure 3.2). According to GMYC, 53 MorphOTUs were incongruent with molecular classification. The same number of MorphOTUs were merged and mixed by ASAP as well as by GMYC. GMYC split the other nine MorphOTUs, considered congruent with ASAP.

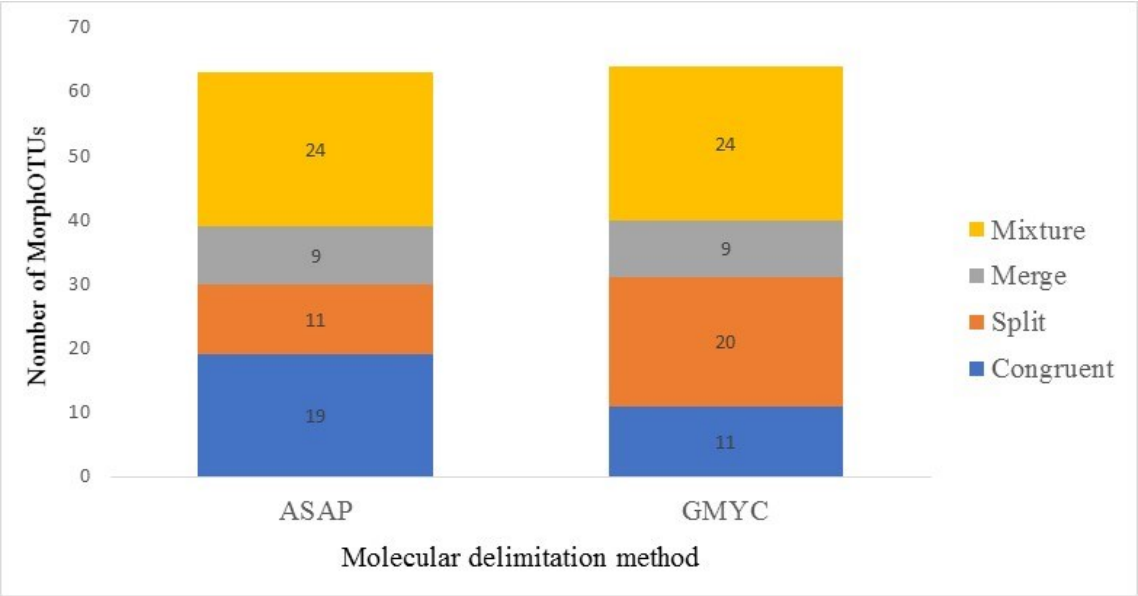


Figure 3.2. Correspondence between MorphOTUs delimited according to morphological assessment and MOTUs delimited with ASAP and GMYC delimitation methods. The graphs show the distribution of the rearrangements in the MorphOTUs, as suggested by the ASAP and GMYC analysis.

There was no split or merge between different super-families of Hymenoptera. Differences between MorphOTUs and MOTUs were within the seven super-families studied.

3.6.2.1.MOTUs delimitation within Hymenoptera super-families

Chalcidoidea was estimated to include 34 different MorphOTUs, however molecular delimitation methods recorded between 31 and 49 MOTUs (Table 3.4). Nine among 19 non-singleton MorphOTUs belonging to Chalcidoidea were congruent with MOTUs based on ASAP (Figure 3.3), however only four were similarly delimited according to GMYC (Figure 3.3). The other 11 MorphOTUs were mainly merged with other MorphOTUs judged different morphologically (60%) or mixed (40%). Just one MorphOTU of Chalcidoidea was split into supplementary MOTUs. GMYC tended to split MorphOTUs (33%) or to mix then combine them (66%). Tree-based methods tended to assign some specimens of one MorphOTU to new MOTU, especially observed with Mymaridae and Trichogrammatidae.

Super-family Ichneumonoidea was estimated to contain 24 MorphOTUs, yet it was assigned 29 to 36 MOTUs. The congruence between molecular and morphological delimitation in this Hymenoptera group was low, 22% and 11% according to ASAP and GMYC, respectively (Figure 3.3). ASAP split two MorphOTUs to two and three different MOTUs. Three were

mixed and one MorphOTU was merged to existing MOTU. GMYC promotes splitting rather than merging within Ichneumonidae and Braconidae (Figure 3.3).

Seventeen MorphOTUs established for Platygastroidea were incongruent with molecular analyses. Delimitation methods conceded between 24 and 40 MOTUs. Five rare MorphOTUs were recorded. Like Ichneumonoidea, the overlap between molecular and morphological OTUs was low. Two MorphOTUs were considered to be the same MOTU and no congruence was observed with GMYC (Figure 3.3). ASAP attributed specimens judged different based on their morphological features to other existing MorphOTUs in 60% of cases. It was observed for both Scelionidae and Platygastridae families. GMYC split six among 12 MorphOTUs in this super-family (Figure 3.3).

Molecular analyses for Ceraphronoidea (with 14 MorphOTUs) returned between 26 and 36 MOTUs. Based on ASAP and GMYC, four of 12 non-singleton MorphOTUs were congruent with molecular delimitation. More than 40% of MorphOTUs were mixed with others (41% and 50% with ASAP and GMYC respectively) (Figure 3.3). Two MorphOTUs were split into new MOTUs by ASAP and one by GMYC (Figure 3.3).

Diaprioidea was delimited to have twice or more MOTUs than the estimated MorphOTUs (Table 3.4). Only two MorphOTUs were delimited similarly with ASAP and ABGD. Six from ten abundant MorphOTUs were mixed by ASAP and four by GMYC. Some specimens from the same MorphOTU were considered in the same cluster as other Diaprioidea (Table S6). Splitting was most observed with GMYC (Figure 3.3).

For Cynipoidea, two MOTUs were estimated. Molecular delimitation doubles the number of MorphOTUs. All delimitation methods split the non-singleton MorphOTU of Cynipoidea into two different MOTUs. The same results were recorded for Vespoidea with the sampled Pompilidae family (TableS5, S6).

Unlike the other super-families, Chalcidoidea was the group with the most important overlap between MorphOTUs and MOTUs according to ASAP (47%), followed by Ceraphronoidea (33%), Ichneumonoidea (22%), Diaprioidea (20%) and Platygastroidea (15%). Based on GMYC, Ceraphronoidea (33%) was more congruent than Chalcidoidea (21%), followed by Diaprioidea (20%) and Ichneumonoidea (11%). Platygastroidea present no overlap between the assigned MorphOTUs and MOTUs.

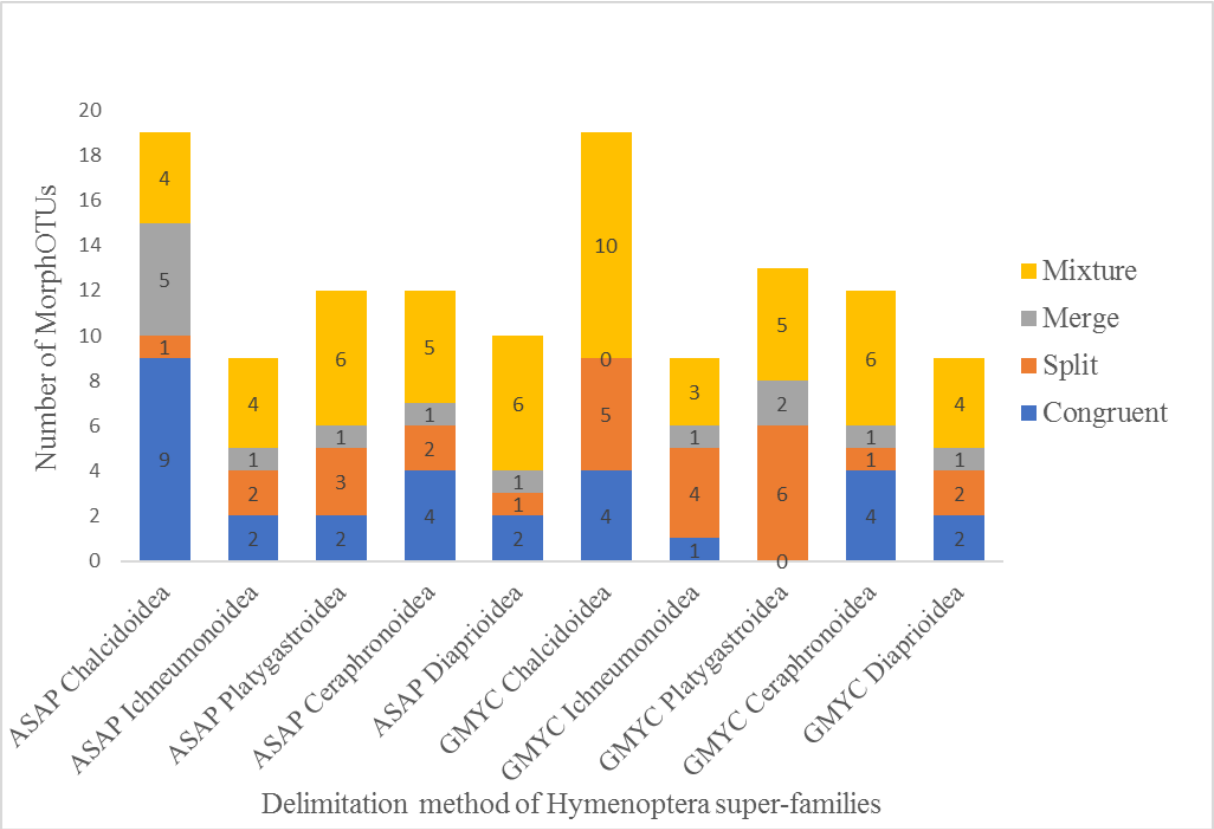


Figure 3.3. Correspondence between MorphOTUs delimited according to morphological assessment and MOTUs delimited with ASAP and GMYC delimitation methods. The graphs show the distribution of the rearrangements in the MorphOTUs, as suggested by the ASAP and GMYC analysis in the most abundant Hymenoptera super-families.

4. Discussion

Our study is the first large scale comparison between morphological and molecular identification in seven abundant and still under-studied super-families of Canadian Hymenoptera. The goal was first to apply a high-throughput DNA barcoding approach to assess its potential to affordably evaluate Hymenoptera diversity. Secondly, we wished to compare the value of molecular and morphological classification. I compared the outcome of parataxonomic sorting with the results of molecular identification and clustering of the same samples.

4.1. High throughput DNA barcoding vs Morphological identification approaches

Our time-cost analyses for molecular and morphological identification showed that the HTS barcoding adopted in this study was approximately half the cost of supplies and time as the morphological identification (Table 3.3). Additionally, the switch from MiSeq to HiSeq

sequencer could reduce the sequencing cost by a factor of five to ten, as reported by Meier *et al*. (2016) and Wang *et al*. (2018).

Ninety-six different combinations were resorted and used to amplify 517 specimens in different plates. We reduced the price from 32CAD per specimen, if uniquely labelled primers were, used to 0.61CAD per specimen. The use of such an approach was safe in our case because every specimen was amplified individually. In addition, we used tags differing by ≥4bp each. We here amplified DNA from specimens separately then sequenced mixed amplicons at low cost (0.44CAD per specimen). We recorded a low cross contamination (2%), unlike standard metabarcoding conducted on bulk samples, which recorded high level of index switching (Piper *et al*., 2019; Creedy *et al*., 2020; Batvaska *et al*., 2021).

We unfortunately lost barcodes from specimens amplified with the tag R1 "caacagtag" during the demultiplexing step. This loss could be due to an experimental biais during Illumina library preparation (Head *et al*., 2014), or it could also be related to the nucleic acid sequence of R1, with respect to Illumina adaptor ligation, leading to sequencing failure or low coverage for most samples (Taub *et al*., 2010). Sanger sequencing these samples may help us understand the reason for failure with this batch of samples.

We failed to obtain 23% and 16% of sequences for Ceraphronoidea and Chalcidoidea samples in reason of lower read number. Working with tiny microhymenoptera can be challenging at all DNA barcoding steps (Spiess *et al*., 2004; Cruaud *et al.*, 2019). We recorded lower intensity bands with ceraphronoids and chalcidoids, which could be caused by the low input DNA quality and quantity or enhancer during PCR amplification such as GC contents (Clarke *et al*., 2017; Krehenwinkel *et al*., 2018; Braukmann *et al*., 2019). Otherwise, the number of reads is shown to be related to the DNA extraction technique. In fact, Carew *et al*. (2018) showed that different extraction methods performed better or worse, depending on taxon, in terms of the number of reads. They obtained >1800 reads when the specimens were homogenized during DNA extraction, but only <20 reads when they were simply immersed (Carew *et al*., 2019). Ceraphronoidea and Chalcidoidea are known to have a poor success rates for DNA barcoding (Cruaud *et al.*, 2019; Ulmer *et al*., 2021; Vasilita *et al*., 2022), leading to a lack of reference DNA barcodes and making these hyperdiverse microhymenoptera groups underdescribed (Bennett *et al*., 2019a; Vasilita *et al*., 2022). We had low DNA barcoding success with eight specimens of Vespoidea (50%). It is difficult to draw conclusions from such a small sampling,

but the low DNA yield may be due to the high level of sclerotization of the specimens, despite their size (Carew *et al*., 2019; Batovska *et al*., 2021).

Most sequences in our study were only identified to family level (82% of sequences), for which reference sequences did not contain genus or species level identification in BOLD or GenBank. Only 72% of our data was assigned to existing BINs in BOLD. Other sequences were not assigned to BINs, a well-documented challenge (Yu *et al*., 2012; Meier *et al*., 2016; Creedy *et al*.,2020; Wang *et al*., 2018; Batovska *et al*., 2021). Only seven sequences belonging to two MOTUs could be confidentially attributed to recognized species. Our results were lower than previously reported for ants (Hymenoptera) (Wang *et al*., 2018), flies (Diptera) (Sonet *et al*., 2013), beetles (Coleoptera) (Pentinsaari *et al*., 2014), and butterflies and moths (Lepidoptera) (Huemer *et al*., 2014), when searching against either or both sequence databases. The unavailability of data for Hymenoptera is probably simply because they are poorly known (Stork *et al*., 2018; Bennett *et al*., 2019a).

Additionally, we performed two alignment methods: *de novo* and assembly to reference sequence. Results of the taxonomic assignment step were incongruent: we found that in some cases, the consensus sequences from assembly to reference gave sequences with more ambiguities, leading to misidentification with the expected species. It should be noted that although there are differences in how *de novo* (Miller et *al.*, 2010) and reference (Gnerre et *al.*, 2009) assembly function, assembly from reference is known to give a generally superior quality of consensus sequence compared to *de novo* assembly (Gnerre et *al.*, 2009). However, this was not the case with our Hymenoptera sequences. As the reference sequence corresponded to a platygastroid, it is logical that the *de novo* assembly outperformed assembly to reference for Ceraphronoidea and Vespoidea groups. However, it was not the case with Ichneumonoidea Chalcidoidea and Diaprioidea; which are phylogenetically closer to Platygastroidea (Peters *et al*., 2017, Figure S8). The BOLD similarity was more informative for an incorrect match (62±20%), however the identification percentage of BOLD was usually high (96±25%). Our results are supported by Meiklejohn *et al*. (2019) who found that BOLD outperformed GenBank. The major problem with BOLD was the inaccessibility of data (3.5% of cases). We provided genus-level identification for 77 sequences that previously had only family-level identifications.

Our approach, which entails HTS barcoding for all assigned MorphOTUs, is not only technically straightforward but also well-suited for quickly creating barcode datasets if

MotphOTUs will be identified at higher levels. It will be especially useful for underdescribed insect groups (Srivathsan *et al*., 2016; Stork *et al*., 2018; Bennett *et al*., 2019a).

4.2.The sources of variance between molecular and morphological delimitation

Our findings highlight the remarkable diversity of Hymenoptera based on both morphological and molecular data (Table 4). The Chalcidoidea were the most abundant and rich insect superfamily (34 MorphOTUs). It is known to be one of the most diverse Hymenoptera groups in the world and in Canada (Rasool *et al*., 2018; Stork *et al*., 2018; Bennett *et al*., 2019a). These numbers were followed by Ichneumonoidea (24 MorphOTUs), Platygastroidea (17 MorphOTUs), Ceraphronoidea (14 MorphOTUs), Diaprioidea (10 MorphOTUs), Cynipoidea (two MorphOTUs) and Vespoidea (one MorphOTU). The dominance of these super-families was also evident in the results of suction trap sampling in 2015 (Favret *et al*., 2019).

If we assume that molecular delimitation provides a good and correct approximation to interspecific divergence in Hymenoptera, then morphological examination previously resolved 51 to 100% in Chalcidoidea, 63 to 82% in Ichneumonoidea, 36 to 70% in Platygastroidea, 36 to 53% in Ceraphronoidea and 35 to 45% in Diaprioidea of the actual diversity. Our results were lower than those recorded by Fernández-Flores *et al*. (2013); Wang *et al*., (2018) and Fagan-Jeffries *et al*. (2018) evaluating Hymenoptera diversity.

About the 75% of specimens were clustered into MOTUs in conflict with MorphOTUs. It is important to remember that the misplacement of one specimen, as a result of a misidentification or contamination of a PCR product, will cause an incongruence; affecting the entire MOTU. Any such possible erroneous placements should be re-evaluated considering additional data (targeting multilocus or identifiction by taxonomists). Yeo *et al*. (2020) reported low congruence for Hymenoptera. The majority of studies examining the levels of congruence between barcodes and morphology concentrate on species-level congruence (Meier *et al*., 2016; Wang *et al*., 2018; Barroso *et al*., 2022). However, as specimens serve as the fundamental building unit of an ecological survey or a museum collection, specimen-level congruence is also crucial. Since biodiversity researchers frequently require abundance and biomass data with species-level resolution (Barwell *et al*., 2015; kemp *et al*., 2017), the proper placement of specimens into species is crucial.

According to research, parataxonomic classification generally yields fewer MorphOTUs than the actual number of species because similar specimens are more frequently lumped than split

(Krell, 2004; Abadie *et al*., 2008; Yeo *et al*., 2020; Parslow *et al*., 2021). Since MorphOTUs are typically neither specified nor given existing names, sorting results are difficult to verify. The inter-subjective falsifiability of the sorting results is challenging, it does not use previously developed biological knowledge, generates typological units, and withholds its sorting criteria; as a result, lumping or splitting MorphOTUs is expected especially in insect groups with a high proportion of cryptic species (Stahlhut *et al*., 2013). Through a comparison of parataxonomic and professional sorting of the same samples Krell (2004) found the latter doubled the number of Hymenoptera OTUs. His research argued that parataxonomy cannot be scientifically trusted, a claim also supported by recent arthropod studies (Saunders, 2018; Shekhovtsov *et al*., 2019; Yeo *et al*., 2020). Our results found many chalcidoid and ceraphronoid MOTUs merged within single MorphOTUs. This effect was probably because most species of these superfamilies are tiny (less than a few millimetres), making it difficult for non-specialists to distinguish between species (Cruaud *et al*., 2018; Rasool *et al*., 2018; Bennett *et al*., 2019a). Taking into account the four molecular delimitation methods, we found numerous cases of OTU splitting within Ichneumonoidea and Platygastroidea. These Hymenoptera groups are especially known to harbour a high degree of species crypsis (Stahlhut *et al*., 2013; Veijalainen *et al*., 2013; Tortorici *et al*., 2019).

We cannot make any meaningful conclusions with respect to rare MorphOTUs, those with fewer than 10 specimens. Other studies demonstrated that for species with few specimens, even those that show no or little intraspecific genetic divergence, species crypsis cannot be detected unless more sequences are added, especially from different sampling habitats (Janzen *et al*., 2005; Wenker *et al*., 2016). Relying only on genetic methods for defining cryptic species could provide inaccurate findings. It requires the addition of other empirical data to be seriously supported.

However, there are many biological reasons why it is unwise to anticipate the splitting of MorphOTUs that are genetically placed in the same cluster. Intraspecific polymorphism in some groups can explain a poor accuracy (Luz *et al*., 2020): developmental changes resulting in various colors or sizes abound in the same species, since parasitoid development is closely related to host quality (Saunders, 2018; Belokobyl'skij *et al*., 2019). These kinds of difficulties could be observed with Platygastroidea, Diaprioidea and Ceraphronoidea molecular delimitation methods tending to lump different MorphOTUs. These superfamilies are known to have great polymorphism within species (Chen, 2018; Luz *et al*., 2020; Ulmer *et al*., 2021).

Among the errors in morphological sorting is to split males from females of the same MorphOTU (Saunders, 2018). It can be the case with Chalcidoidea (Mymaridae), numerous MorphOTUs were lumped together based on molecular delimitation methods. In fact, this family is known to exhibit strong sexual dimorphism, especially with regard to flagellum segment number and shape (Mockford, 1997; Saleem and Anis, 2021). A taxonomic expertise may be required for dissection and slide-mounting of representative specimens (Ulmer *et al*., 2021; Vasilita *et al*., 2022).

Based on the placement of the individual samples in the microtiter plates, cross contamination of specimens during DNA extraction or PCR preparation may also contribute to the inconsistency between MOTUs and MorphOTUs. It could be one reason explaining mixed MorphOTUs, where an individual of one MorphOTU was lumped with another individual from an otherwise clearly different MorphOTU. It was mostly observed with very small Chalcidoidea. Their minuscule size makes their manipulation challenging, whether for morphological or molecular analysis.

As few as 37% of MorphOTUs were represented by singletons and DNA species delimitation methods are known to be affected by restricted sampling effort (Ahrens *et al*., 2016). A robust calculation of species limits determined by intra- vs. interspecific variation require sufficient sampling to identify an accurate genetic gap between and within species (Fujisawa and Barraclough 2013). GMYC is especially sensitive to variable species abundance and rarity of species (Fujisawa and Barraclough 2013; Ahrens *et al*., 2016; Luo *et al*., 2018).

Our study is the first based on data sets including morphology and DNA sequencing of every specimen for seven different groups of Hymenoptera. The advantage of our approach is the ability to re-examine the conflicting specimens that were initially misplaced based on morphology and the proportion that were initially misplaced based on DNA barcodes (pre-sorting mistakes). Morphological identification of specimens for which there is a disagreement among the data can be processed by specialists for dissection and slide mounting, in order to ensure the correct MOTU placement.

Using both types of data is more accurate and effective according to “the integrative taxonomy” principles; a taxonomic approach based on the variation of different kinds of data, using as many methods as are feasible (morphology, molecular methods, geography, biology…) (Schlick-Steiner *et al*., 2010; Goulding and Dayrat 2016; Vinarski, 2020). Reaching this level, maybe it will be less accurate to judge the effectiveness of one method compared to the other

(morphological or molecular identification), since our results reported an important conflict between MorphOTUs and MOTUs. It will be more interesting to study the taxonomic and genetic diversity of hundreds of Hymenoptera in order to answer the main question, can MorphOTUs be used as a robust estimator of species diversity (α-diversity and β-diversity)? It will also be important to compare the different contributions of genetic diversity to the morphological diversity by network analysis to promote biodiversity assessment improvement; in order to display auxiliary information associated with haplotype sequences, such as the geographic and temporal points at which the samples were collected.

5. Supplementary material

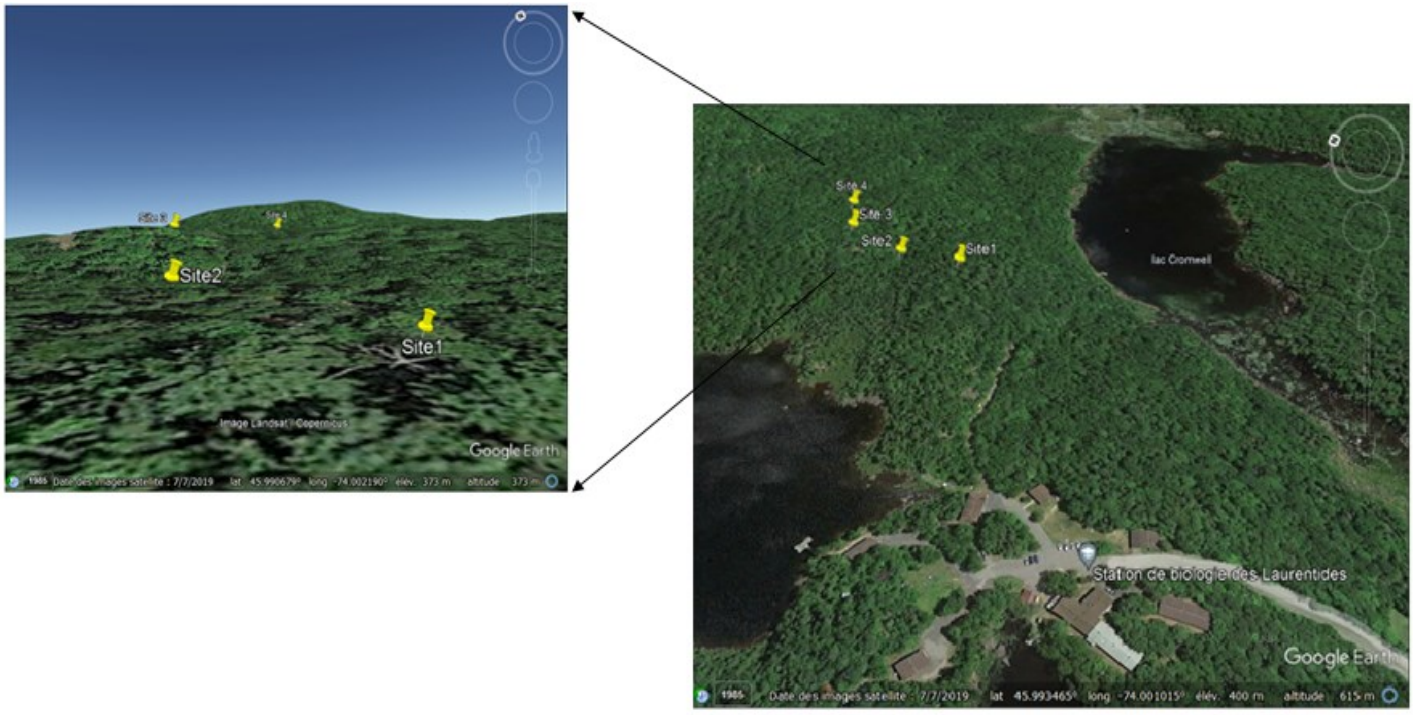


Figure S2. Map of sampled sites localization in the Laurentian forest (Google earth Pro, 2021)

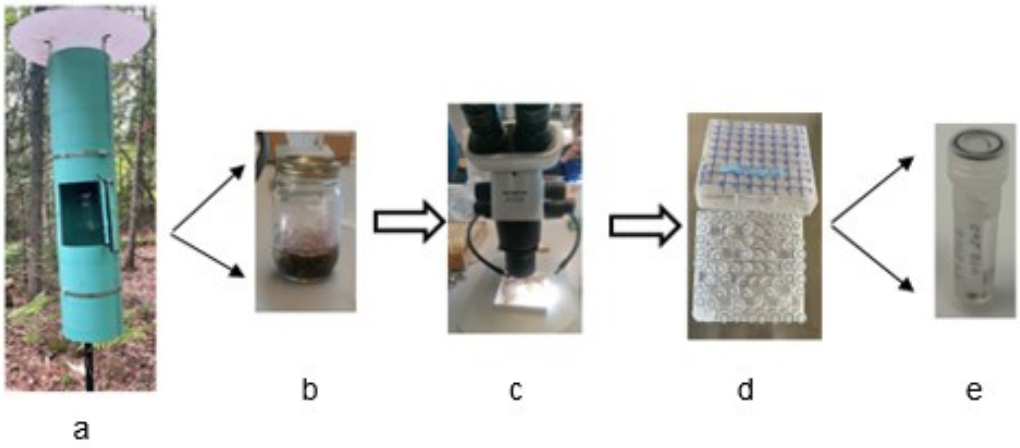


Figure S3. The sorting process of sampled specimens: a: suction trap, b: bottle with the material, specimen sorting under microscope (0.63X), d: Box with different MorphOTUs, e: one MorphOTU.

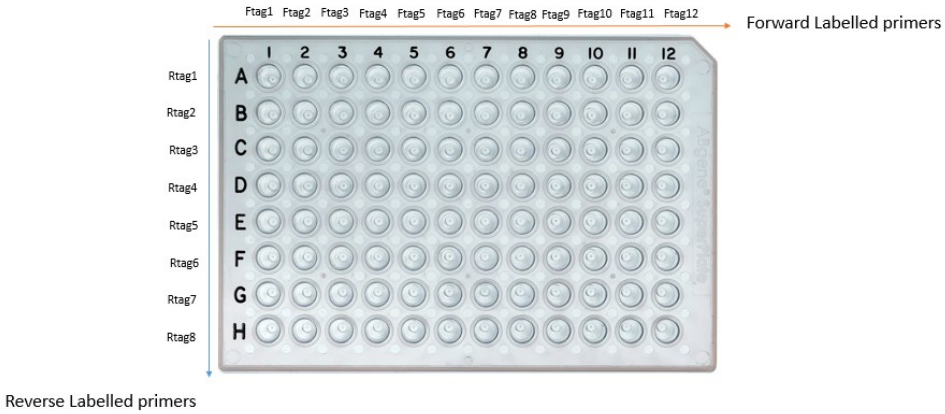


Figure S4. Plaque design using the combination of labelled forward and reverse primers

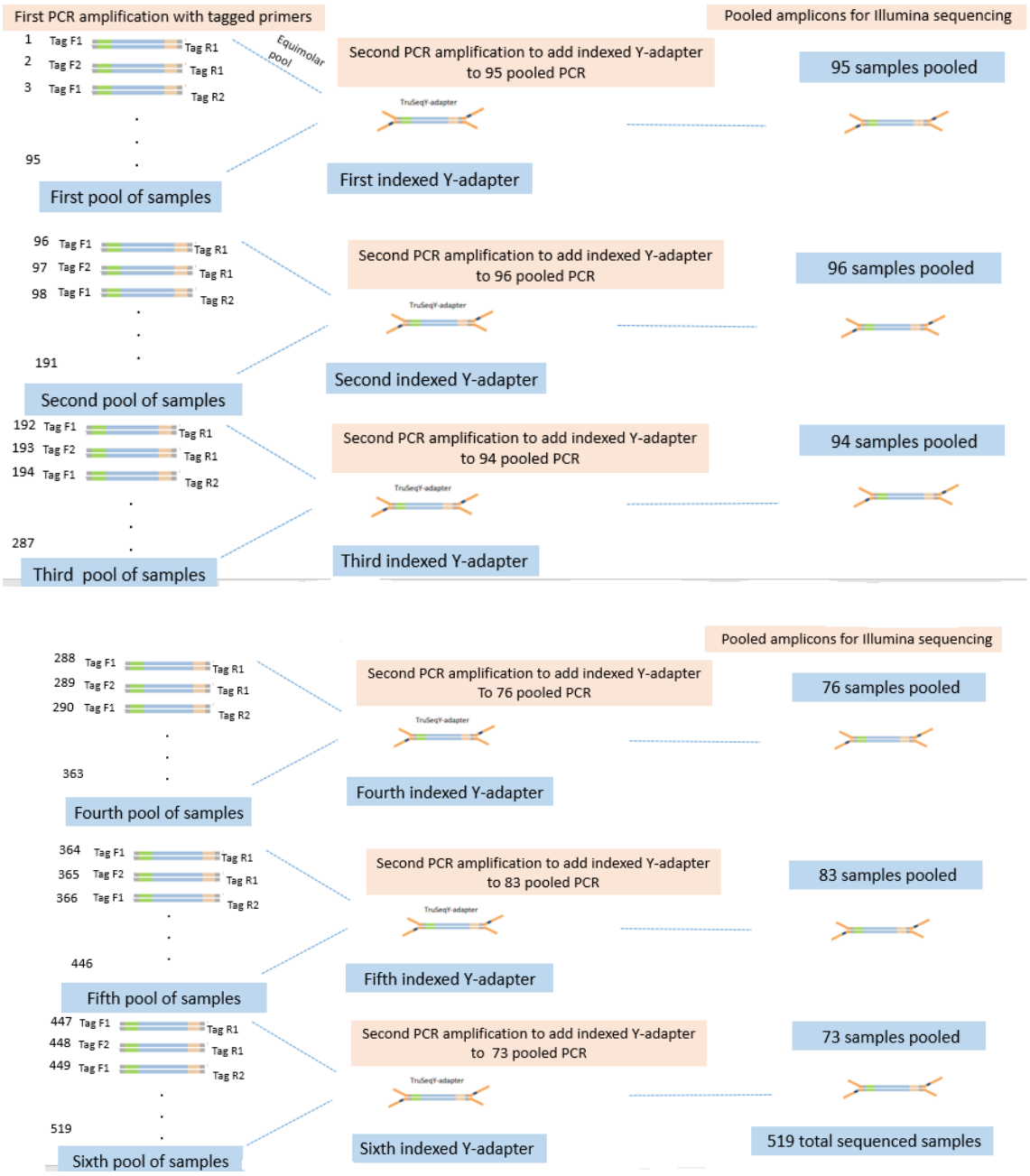


Figure S5. Illustration of the pooling strategy used for library preparation for this research

Nb of species	asap-score	P-val (rank)	W (rank)	Treshold dist.	Text
137	9.00	4.57e-02 (7)	7.14e-05 (11)	0.050676	list csv
* 138	12.00	8.46e-02 (11)	7.08e-05 (13)	0.049165	list csv
125	14.50	6.88e-02 (9)	6.38e-05 (20)	0.059457	list csv
* 140	15.00	9.51e-02 (12)	6.78e-05 (18)	0.048728	list csv
* 148	21.00	1.15e-02 (6)	5.55e-05 (36)	0.043299	list csv
* 158	23.50	4.83e-03 (3)	5.31e-05 (44)	0.035540	list csv
119	30.00	6.15e-03 (4)	4.94e-05 (56)	0.063473	list csv
* 147	30.00	1.96e-01 (23)	5.54e-05 (37)	0.043838	list csv
* 154	30.00	2.18e-01 (27)	5.66e-05 (33)	0.038219	list csv
121	30.50	6.92e-02 (10)	5.13e-05 (51)	0.061911	list csv

Figure S6. A list of the "best" partitions (10 by default) generated by ASAP

```
/*
/* ABGD (Automatic Barcod Gap Discovery)
/* web version 08/26/21 - 09:51AM

Fasta Format detected
Nb seq:443 */
Partition 1 : found 178 groups (prior maximal distance P= 0.005000)
Partition 2 : found 178 groups (prior maximal distance P= 0.005854)
Partition 3 : found 174 groups (prior maximal distance P= 0.006854)
Partition 4 : found 174 groups (prior maximal distance P= 0.008024)
Partition 5 : found 174 groups (prior maximal distance P= 0.009394)
Partition 6 : found 174 groups (prior maximal distance P= 0.010999)
Partition 7 : found 154 groups (prior maximal distance P= 0.012877)
Partition 8 : found 151 groups (prior maximal distance P= 0.015076)
Partition 9 : found 151 groups (prior maximal distance P= 0.017651)
Partition 10 : found 1 groups (prior maximal distance P= 0.020666)
```

Figure S7. A list of the "best" partitions (10 by default) generated by ABGD

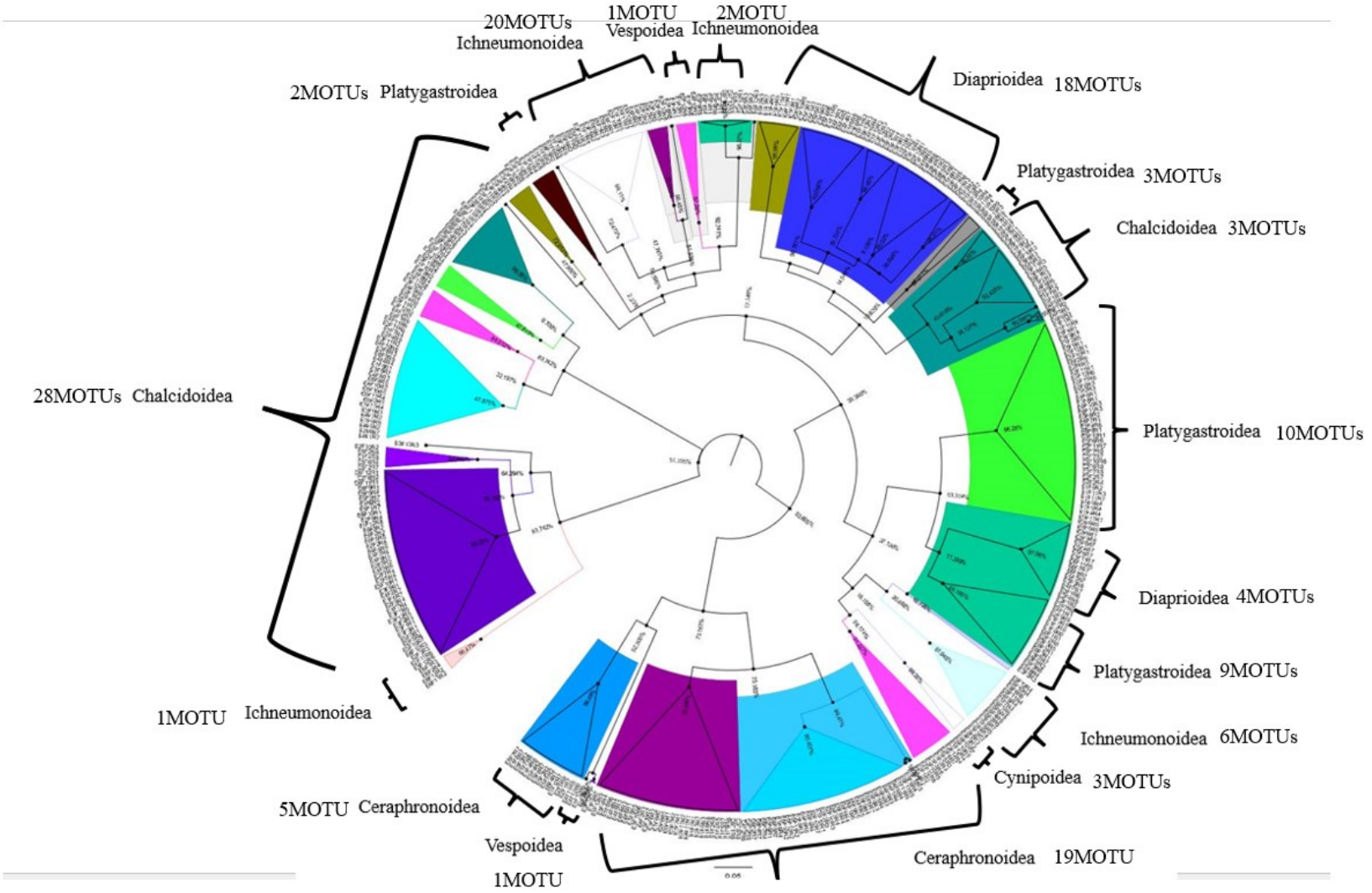


Figure S8. Phylogram of Hymenoptera super-families obtained by BEAST and corresponding MOTUs number as delimited by ASAP

Table S3. List of tailed primers used to amplify Hymenoptera

Primer name	Tailed primer sequence (5′–3′)
TagF1	aacagatgg
TagF2	accgattcg
TagF3	acgaagtga
TagF4	acgtgtcag
TagF5	actatgccg
TagF6	actcatcgt
TagF7	actgccata
TagF8	aggactaca
TagF9	agtacctag
TagF10	agtgaagca
TagF11	atatgcacg
TagF12	atcgcacga
TagR1	caacagtag
TagR2	caccgtcat
TagR3	cactgctta
TagR4	cagaaccaa
TagR5	cagacgatg
TagR6	cagctagca
TagR7	cctaacact
TagR8	cgcaaggat

Table S4. Evolutionary models suggested by PhyML software

Model	Decoration	K	LIK	AIC	BIC
GTR	+R	900	-25788,21499	53376,42998	57628,96025
GTR	+G+I	896	-25812,06774	53416,13548	57649,76562
GTR	+G	895	-25865,15215	53520,3043	57749,2094
TN93	+R	897	-25923,01824	53640,03648	57878,39165
GTR	+I	895	-27107,43437	56004,86874	60233,77384
GTR		894	-28085,03098	57958,06196	62182,24203

Table S5. MOTUs Delimitation methods results

MorphOTUs	Number of specimens	ASAP MOTU number	ABGD MOTU number	GMYC MOTU number	bPTTP MOTU number	Super- family
MorphOTU 2	3	2	2	2	2	Diaprioidea
MorphOTU 5	2	1	1	1	1	Diaprioidea
MorphOTU 8	4	2	2	2	2	Ceraphronoidae
MorphOTU 10	2	2	2	2	2	Ceraphronoidea
MorphOTU 13	4	2	2	2	2	Platygastroidea
MorphOTU 14	5	3	3	3	3	Diaprioidea
MorphOTU 21	14	10	10	12	11	Diaprioidea
MorphOTU 24	14	4	4	7	9	Chalcidoidea
MorphOTU 28	10	4	5	6	6	Platygastroidea
MorphOTU29	9	1	1	5	2	Chalcidoidea
MorphOTU30	3	1	1	1	1	Ceraphronoidea
MorphOTU 34	1	1	1	1	1	Ichneumonoidea
MorphOTU36	2	2	2	2	2	Platygastroidea
MorphOTU42	1	1	1	1	1	Platygastroidea
MorphOTU43	1	1	1	1	1	Chalcidoidea
MorphOTU45	3	2	2	2	2	Chalcidoidea
MorphOTU46	5	3	3	4	4	Platygastroidea
MorphOTU48	4	2	3	3	3	Chalcidoidea
MorphOTU51	1	1	1	1	1	Ceraphronidea
MorphOTU53	6	4	4	6	4	Ichneumonoidea
MorphOTU56	1	1	1	1	1	Ichneumonoidea
MorphOTU58	1	1	1	1	1	Chalcidoidea
MorphOTU63	12	5	6	7	6	Platygastroidea
MorphOTU65	1	1	1	1	1	Platygastroidea
MorphOTU66	7	4	4	5	4	Ichneumonoidea

MorphOTU68	2	1	1	1	1	Ceraphronoidea
MorphOTU80	1	1	1	1	1	Chalcidoidea
MorphOTU81	14	4	4	4	4	Ceraphronoidea
MorphOTU86	8	3	3	3	3	Ichneumonoidea
MorphOTU88	1	1	1	1	1	Ichneumonoidea
MorphOTU92	1	1	1	1	1	Chalcidoidea
MorphOTU102	4	2		2	2	Chalcidoidea
MorphOTU103	2	2	2	2	2	Platygastroidea
MorphOTU104	3	1	1	1	1	Ceraphronoidea
MorphOTU113	1	1	1	1	1	Ichneumonoidea
MorphOTU119	7	3	3	3	3	Ceraphronoidea
MorphOTU122	7	3	3	3	3	Diaprioidea
MorphOTU125	4	2	2	2	2	Chalcidoidea
MorphOTU126	1	1	1	1	1	Chalcidoidea
MorphOTU127	13	2	2	4	2	Platygastroidea
MorphOTU131	14	2	2	5	2	Chalcidoidea
MorphOTU134	7	5	5	5	6	Diaprioidea
MorphOTU135	2	2	2	2	2	Diaprioidea
MorphOTU136	1	1	1	1	1	Diaprioidea
MorphOTU141	1	1	1	1	1	Ichneumonoidea
MorphOTU142	1	1	1	1	1	Ichneumonoidea
MorphOTU144	2	2	2	2	2	Chalcidoidea
MorphOTU146	2	1	1	1	1	Ceraphronoidea
MorphOTU150	5	4	4	4	4	Ceraphronoidea
MorphOTU154	2	2	2	2	2	Chalcidoidea
MorphOTU157	7	3	3	4	4	Chalcidoidea
MorphOTU159	4	2	2	3	2	Pompilidae
MorphOTU160	18	4	4	9	5	Chalcidoidea
MorphOTU168	1	1	1	1	1	Ichneumonoidea

MorphOTU171	1	1	1	1	1	Chalcidoidea
MorphOTU175	1	1	1	1	1	Ichneumonoidea
MorphOTU176	6	1	1	4	1	Chalcidoidea
MorphOTU177	4	1	1	1	1	Diaproidea
MorphOTU179	1	1	1	1	1	Cynipoidea
MorphOTU183	2	1	2	2	2	Platygastroidea
MorphOTU185	5	2	2	2	2	Ceraphronoidea
MorphOTU186	3	2	2	2	2	Platygastroidea
MorphOTU187	2	1	1	2	1	Chalcidoidea
MorphOTU188	13	1	1	5	3	Chalcidoidea
MorphOTU191	1	1	1	1	1	Ichneimonoidea
MorphOTU192	2	1	1	1	1	Ichneumonoidea
MorphOTU193	1	1	1	1	1	Chalcidoidea
MorphOTU194	2	2	2	2	2	Ichneumonoidea
MorphOTU195	1	1	1	1	1	Chalcidoidea
MorphOTU196	1	1	1	1	1	Chalcidoidea
MorphOTU199	16	4	3	4	7	Ichneumonoidea
MorphOTU200	3	1	1	3	1	Platygastroidea
MorphOTU201	31	4	5	12	11	Ceraphronoidea
MorphOTU202	21	5	5	9	9	Platygastroidea
MorphOTU203	3	2	2	2	2	Cynipoidea
MorphOTU204	1	1	1	1	1	Ichneumonoidea
MorphOTU205	3	2	2	2	2	Platygastroidea
MorphOTU206	1	1	1	1	1	Chalcidoidea
MorphOTU207	1	1	1	1	1	Chalcidoidea
MorphOTU209	1	1	1	1	1	Chalcidoidea
MorphOTU210	2	1	2	2	2	Chalcidoidea
MorphOTU211	2	1	1	1	1	Chalcidoidea
MorphOTU212	14	3	3	3	3	Ichneumonoidea

MorphOTU213	5	3	3	4	3	Platygastroidea
MorphOTU214	7	2	2	4	3	Chalcidoidea
MorphOTU215	9	7	7	7	8	Ceraphronoidea
MorphOTU216	1	1	1	1	1	Ichneumonoidea
MorphOTU217	1	1	1	1	1	Ichneumonoidea
MorphOTU218	1	1	1	1	1	Ceraphronoidea
MorphOTU220	6	4	4	4	4	Diaprioidea
MorphOTU221	1	1	1	1	1	Chalcidoidea
MorphOTU222	2	2	2	2	2	Ichneumonoidea
MorphOTU223	2	1	1	1	1	Chalcidoidea
MorphOTU224	1	1	1	1	1	Platygastroidea
MorphOTU225	1	1	1	1	1	Chalcidoidea
MorphOTU226	2	1	2	2	2	Ichneumonoidea
MorphOTU228	1	1	1	1	1	Ichneumonoidea
MorphOTU229	1	1	1	1	1	Ichneumonoidea
MorphOTU234	1	1	1	1	1	Ichneumonoidea
MorphOTU235	1	1	1	1	1	Chalcidoidea
MorphOTU236	2	1	1	1	1	Chalcidoidea
MorphOTU238	1	1	1	1	1	Platygastroidea

Table S6. MOTUs delimitation results; Morphological vs molecular delimitation. The numbers in the table correspond to the MOTU number delimited by each method

Sequences ID	MorphOTUs	ASAP	ABGD	GMYC	bPTP	Singleton MOTUs	Super-families
E1F5R6	2	24	23	75	98		Diaprioidea
E3F4R8	2	93	99	84	112		Diaprioidea
E6F5R3	2	93	99	84	112		Diaprioidea
E2F8R1	5	23	77	35	150		Diaprioidea
E5F5R6	5	23	77	35	150		Diaprioidea
E1F6R3	8	28	27	6	82		Ceraphronoidae
E2F3R7	8	28	27	6	82		Ceraphronoidae
E5F1R1	8	28	27	6	82		Ceraphronoidae
E3F5R8	8	95	101	102	144		Ceraphronoidae
E1F12R6	10	57	57	107	1		Ceraphronoidae
E2F10R3	10	41	41	106	188		Ceraphronoidae
E1F3R8	13	15	14	145	24		Platygastroidea
E1F5R3	13	21	20	39	25		Platygastroidea
E2F12R4	13	21	20	39	25		Platygastroidea
E2F6R7	13	21	20	39	25		Platygastroidea
E1F3R7	14	14	13	73	104		Diaprioidea
E3F4R5	14	14	13	73	104		Diaprioidea
E3F9R6	14	14	13	73	104		Diaprioidea
E6F2R3	14	14	13	73	104		Diaprioidea
E1F10R6	14	6	5	203	105		Diaprioidea
E2F1R3	21	60	60	214	27		Diaprioidea
E2F4R4	21	24	67	134	33		Diaprioidea
E2F4R3	21	64	66	211	71		Diaprioidea
E1F10R1	21	45	45	81	73		Diaprioidea

E2F12R1	21	45	45	81	73		Diaprioidea
E2F7R2	21	45	45	81	73		Diaprioidea
E3F6R8	21	24	23	75	98		Diaprioidea
E1F11R5	21	53	53	205	99		Diaprioidea
E5F6R6	21	53	53	206	100		Diaprioidea
E1F10R4	21	47	47	210	103		Diaprioidea
E1F10R3	21	46	46	208	132		Diaprioidea
E5F9R5	21	53	53	78	133		Diaprioidea
E2F12R3	21	53	53	79	133		Diaprioidea
E4F3R2	21	117	128	162	164		Diaprioidea
E4F1R7	24	114	125	193	85		Chalcidoidea
E4F4R4	24	23	39	197	86		Chalcidoidea
E4F8R3	24	23	39	66	87		Chalcidoidea
E1F9R1	24	23	39	66	88		Chalcidoidea
E3F9R7	24	23	39	66	89		Chalcidoidea
E4F7R2	24	44	44	64	90		Chalcidoidea
E1F11R4	24	23	39	67	91		Chalcidoidea
E4F10R2	24	23	39	67	91		Chalcidoidea
E6F10R5	24	23	39	67	91		Chalcidoidea
E6F2R5	24	23	39	67	91		Chalcidoidea
E6F8R5	24	23	39	67	91		Chalcidoidea
E6F11R4	24	23	39	67	92		Chalcidoidea
E6F3R5	24	23	39	195	95		Chalcidoidea
E2F6R8	24	69	72	110	170		Chalcidoidea
E6F1R1	28	40	40	123	64		Platygastroidea
E3F3R5	28	40	40	34	66		Platygastroidea
E6F2R1	28	40	40	34	66		Platygastroidea
E6F3R1	28	40	40	34	66		Platygastroidea
E1F1R7	28	4	4	118	130		Platygastroidea

E2F7R7	28	72	75	111	174		Platygastroidea
E3F12R1	28	84	103	28	177		Platygastroidea
E3F6R6	28	84	103	28	177		Platygastroidea
E2F11R7	28	84	90	27	178		Platygastroidea
E5F7R5	28	84	90	27	178		Platygastroidea
E3F11R5	29	17	16	201	175		Chalcidoidea
E6F3R3	29	17	16	69	176		Chalcidoidea
E6F5R2	29	17	16	69	176		Chalcidoidea
E6F7R2	29	17	16	69	176		Chalcidoidea
E3F1R5	29	17	16	70	176		Chalcidoidea
E3F5R6	29	17	16	71	176		Chalcidoidea
E6F4R2	29	17	16	71	176		Chalcidoidea
E6F6R2	29	17	16	71	176		Chalcidoidea
E3F10R4	29	17	16	200	176		Chalcidoidea
E1F5R7	30	25	24	5	145		Ceraphronidea
E1F9R6	30	25	24	5	145		Ceraphronidea
E6F1R6	30	25	24	5	145		Ceraphronidea
E4F6R3	34	123	134	170	17	S	Ichneumonidea
E4F11R3	36	129	140	171	45		Platygastroidea
E3F9R8	36	101	112	120	160		Platygastroidea
E1F2R6	42	2	2	141	94	S	Platygastroidea
E2F11R8	43	85	91	183	143	S	Chalcidoidea
E1F1R2	45	1	1	59	139		Chalcidoidea
E5F7R1	45	1	1	59	139		Chalcidoidea
E2F7R4	45	26	25	216	166		Chalcidoidea
E1F3R5	46	13	12	33	57		Platygastroidea
E2F2R7	46	40	40	121	146		Platygastroidea
E2F7R5	46	40	40	29	147		Platygastroidea
E1F11R8	46	55	55	31	161		Platygastroidea

E5F8R1	46	55	55	31	161		Platygastroidea
E1F9R8	48	44	44	64	90		Chalcidoidea
E2F8R3	48	23	78	192	96		Chalcidoidea
E2F8R7	48	23	80	68	162		Chalcidoidea
E4F6R2	48	23	80	68	162		Chalcidoidea
E4F3R3	51	118	129	100	21	S	Ceraphronoidea
E1F5R5	53	86	22	160	35		Ichneumonoidea
E6F4R5	53	136	149	151	41		Ichneumonoidea
E4F7R5	53	106	117	152	43		Ichneumonoidea
E3F10R8	53	106	117	153	43		Ichneumonoidea
E4F2R4	53	106	117	154	43		Ichneumonoidea
E2F12R5	53	86	92	155	44		Ichneumonoidea
E2F5R7	56	56	56	57	53	S	Ichneumonoidae
E3F11R8	58	111	122	181	38	S	Chalcidoidea
E2F9R8	63	78	84	136	26		Platygastroidea
E2F2R6	63	40	40	122	65		Platygastroidea
E2F7R8	63	40	76	32	129		Platygastroidea
E4F1R1	63	40	76	32	129		Platygastroidea
E6F6R3	63	137	150	119	131		Platygastroidea
E1F2R5	63	8	7	25	171		Platygastroidea
E5F11R1	63	8	7	25	171		Platygastroidea
E5F10R1	63	8	7	26	171		Platygastroidea
E5F12R1	63	8	7	26	171		Platygastroidea
E5F9R1	63	8	7	26	171		Platygastroidea
E6F2R6	63	8	7	26	171		Platygastroidea
E3F1R6	63	8	65	129	194		Platygastroidea
E1F3R4	65	12	11	146	23	S	Platygastroidea
E1F7R8	66	34	33	51	5		Ichneumonoidea
E2F11R3	66	34	33	51	5		Ichneumonoidea

E5F3R6	66	34	33	164	5		Ichneumonoidea
E1F12R3	66	56	56	163	6		Ichneumonoidea
E1F10R8	66	50	50	163	6		Ichneumonoidea
E1F10R5	66	48	48	48	7		Ichneumonoidea
E2F10R1	66	48	48	48	7		Ichneumonoidea
E1F7R1	68	30	29	8	81		Ceraphronoidae
E5F1R2	68	30	29	8	81		Ceraphronoidae
E5F8R6	80	134	147	177	138	S	Chalcidoidea
E2F12R7	81	57	57	20	1		Ceraphronoidae
E1F12R7	81	58	58	108	155		Ceraphronoidae
E1F4R6	81	19	18	23	156		Ceraphronoidae
E1F7R6	81	19	18	23	156		Ceraphronoidae
E1F8R7	81	19	18	23	156		Ceraphronoidae
E5F2R2	81	19	18	23	156		Ceraphronoidae
E5F3R2	81	19	18	23	156		Ceraphronoidae
E5F4R2	81	19	18	23	156		Ceraphronoidae
E5F5R2	81	19	18	23	156		Ceraphronoidae
E5F7R7	81	19	18	23	156		Ceraphronoidae
E5F8R7	81	19	18	23	156		Ceraphronoidae
E6F4R1	81	19	18	23	156		Ceraphronoidae
E2F1R7	81	62	62	21	182		Ceraphronoidae
E2F5R8	81	62	62	21	182		Ceraphronoidae
E4F4R1	86	119	130	49	4		Ichneumonoidea
E6F7R3	86	119	130	49	4		Ichneumonoidea
E6F8R3	86	119	130	49	4		Ichneumonoidea
E6F9R3	86	119	130	49	4		Ichneumonoidea
E5F6R2	86	34	33	50	5		Ichneumonoidea
E5F7R2	86	34	33	50	5		Ichneumonoidea
E5F8R2	86	34	33	50	5		Ichneumonoidea

E1F6R4	86	29	28	165	8		Ichneumonoidea
E1F8R2	88	35	34	142	93	S	Ichneumonoidae
E2F4R5	92	65	68	190	83	S	Chalcidoidea
E4F11R6	102	110	121	63	84		Chalcidoidea
E4F1R6	102	110	121	63	84		Chalcidoidea
E4F3R7	102	110	121	63	84		Chalcidoidea
E3F1R2	102	23	121	132	151		Chalcidoidea
E3F3R2	103	40	40	128	68		Platygastroidea
E1F2R4	103	2	2	38	94		Platygastroidea
E3F5R5	104	94	100	18	47		Ceraphronoidae
E6F5R1	104	94	100	18	47		Ceraphronoidae
E6F8R2	104	94	100	18	47		Ceraphronoidae
E3F5R2	113	34	33	51	34	S	Ichneumonidae
E1F9R3	119	41	41	12	50		Ceraphronoidae
E5F10R2	119	41	41	12	50		Ceraphronoidae
E5F11R2	119	41	41	12	50		Ceraphronoidae
E5F12R2	119	41	41	12	50		Ceraphronoidae
E2F8R5	119	74	79	4	77		Ceraphronoidae
E6F6R1	119	74	79	4	77		Ceraphronoidae
E3F6R2	119	38	37	2	121		Ceraphronoidae
E5F1R3	122	53	143	76	28		Diaprioidea
E5F2R3	122	53	143	76	28		Diaprioidea
E5F5R3	122	53	143	76	28		Diaprioidea
E5F3R3	122	132	145	77	30		Diaprioidea
E5F4R3	122	132	145	77	30		Diaprioidea
E1F6R1	122	27	26	83	113		Diaprioidea
E2F1R5	122	27	26	83	113		Diaprioidea
E3F7R2	125	23	106	55	15		Chalcidoidea
E4F5R1	125	23	106	55	15		Chalcidoidea

E1F10R7	125	49	49	62	137		Chalcidoidea
E4F8R2	125	49	49	62	137		Chalcidoidea
E3F11R6	126	110	121	63	84	S	Chalcidoidea
E1F2R3	127	8	7	24	171		Platygastroidea
E2F4R8	127	8	7	24	171		Platygastroidea
E3F8R2	127	8	7	24	171		Platygastroidea
E4F6R1	127	8	7	24	171		Platygastroidea
E6F10R3	127	8	7	24	171		Platygastroidea
E6F11R3	127	8	7	24	171		Platygastroidea
E6F7R1	127	8	7	24	171		Platygastroidea
E6F8R1	127	8	7	24	171		Platygastroidea
E5F10R6	127	8	7	24	171		Platygastroidea
E2F6R5	127	8	7	25	171		Platygastroidea
E5F11R6	127	8	7	26	171		Platygastroidea
E5F12R6	127	8	7	26	171		Platygastroidea
E5F6R3	127	8	7	26	171		Platygastroidea
E1F2R7	131	9	8	56	40		Chalcidoidea
E2F11R4	131	9	8	56	40		Chalcidoidea
E5F10R3	131	9	8	56	40		Chalcidoidea
E5F11R3	131	9	8	56	40		Chalcidoidea
E5F7R3	131	9	8	56	40		Chalcidoidea
E5F8R3	131	9	8	56	40		Chalcidoidea
E5F8R5	131	9	8	56	40		Chalcidoidea
E5F9R3	131	9	8	56	40		Chalcidoidea
E1F7R7	131	9	8	56	40		Chalcidoidea
E2F5R2	131	9	8	56	40		Chalcidoidea
E4F8R6	131	9	8	56	40		Chalcidoidea
E5F1R7	131	9	8	174	40		Chalcidoidea
E1F5R2	131	9	8	175	40		Chalcidoidea

E3F5R4	131	17	16	70	176		Chalcidoidea
E4F4R7	134	121	132	72	97		Diaproidae
E1F2R2	134	7	6	80	102		Diaproidae
E5F12R3	134	51	51	72	107		Diaproidae
E1F11R2	134	51	51	72	108		Diaproidae
E4F7R6	134	125	136	213	114		Diaprioidea
E5F1R4	134	51	51	78	133		Diaproidae
E5F2R4	134	51	51	78	133		Diaproidae
E5F3R4	135	7	6	80	102		Diaprioidea
E3F9R2	135	51	51	78	133		Diaprioidea
E4F10R4	136	128	139	202	109	S	Diaprioidea
E4F9R4	141	127	138	157	115	S	Ichneumonoidea
E1F4R1	142	16	15	169	2	S	Ichneumonoidea
E3F4R4	144	91	97	179	16		Chalcidoidea
E2F3R3	144	26	64	221	111		Chalcidoidea
E1F12R5	146	43	43	1	10		Ceraphronoidea
E1F9R7	146	43	43	1	10		Ceraphronoidea
E5F4R4	150	133	146	99	9		Ceraphronoidae
E3F10R2	150	102	113	101	120		Ceraphronoidae
E1F8R5	150	38	37	2	121		Ceraphronoidae
E1F12R4	150	19	18	23	156		Ceraphronoidea
E5F10R5	150	19	18	23	156		Ceraphronoidea
E2F10R4	154	79	85	184	134		Chalcidoidea
E3F11R2	154	107	118	186	136		Chalcidoidea
E4F4R5	157	91	97	60	16		Chalcidoidea
E1F4R8	157	20	19	95	79		Chalcidoidea
E2F10R2	157	20	19	95	79		Chalcidoidea
E5F2R6	157	20	19	95	79		Chalcidoidea
E3F12R6	157	26	25	219	157		Chalcidoidea

E1F6R2	157	26	25	89	166		Chalcidoidea
E5F5R4	157	26	25	89	166		Chalcidoidea
E4F12R2	159	130	142	168	18		Vespoidea
E1F3R1	159	3	3	53	19		Vespoidea
E2F12R6	159	3	3	53	19		Vespoidea
E1F1R6	159	3	3	167	19		Vespoidea
E5F2R7	160	131	144	96	78		Chalcidoidea
E5F6R4	160	131	144	96	78		Chalcidoidea
E2F11R6	160	83	89	222	110		Chalcidoidea
E3F1R7	160	26	25	89	166		Chalcidoidea
E3F2R7	160	26	25	89	166		Chalcidoidea
E6F10R1	160	26	25	89	166		Chalcidoidea
E6F12R1	160	26	25	89	166		Chalcidoidea
E6F9R1	160	26	25	89	166		Chalcidoidea
E2F6R2	160	26	25	90	166		Chalcidoidea
E6F11R1	160	26	25	90	166		Chalcidoidea
E1F12R8	160	26	25	92	166		Chalcidoidea
E2F9R4	160	26	25	92	166		Chalcidoidea
E4F2R7	160	26	25	217	166		Chalcidoidea
E2F2R8	160	26	25	218	166		Chalcidoidea
E2F5R4	160	26	25	86	169		Chalcidoidea
E3F10R7	160	26	25	86	169		Chalcidoidea
E4F12R3	160	26	25	86	169		Chalcidoidea
E1F8R1	160	26	25	86	169		Chalcidoidea
E4F8R1	168	96	104	44	116	S	Ichneumonoidea
E3F1R3	171	88	94	173	14	S	Chalcidoidea
E2F10R8	175	81	87	137	76	S	Ichneumonoidae
E1F4R2	176	17	16	69	176		Chalcidoidea
E6F12R4	176	17	16	69	176		Chalcidoidea

E6F1R2	176	17	16	69	176		Chalcidoidea
E4F11R2	176	17	16	70	176		Chalcidoidea
E3F2R3	176	17	16	198	176		Chalcidoidea
E5F7R4	176	17	16	199	176		Chalcidoidea
E1F11R1	177	5	5	74	106		Diaprioidea
E1F2R1	177	5	5	74	106		Diaprioidea
E2F8R2	177	5	5	74	106		Diaprioidea
E4F10R1	177	5	5	74	106		Diaprioidea
E1F3R2	179	11	10	149	70	S	Cynipoidea
E3F6R4	183	40	102	124	148		Platygastroidea
E3F3R3	183	40	40	124	149		Platygastroidea
E1F2R8	185	10	9	7	59		Ceraphronidea
E1F4R5	185	10	9	7	59		Ceraphronidea
E1F8R8	185	10	9	7	59		Ceraphronoidea
E2F11R2	185	82	88	9	80		Ceraphronoidea
E4F11R1	185	82	88	9	80		Ceraphronoidea
E1F3R6	186	13	12	33	57		Platygastroidea
E1F6R6	186	13	12	33	57		Platygastroidea
E3F3R6	186	40	40	126	69		Platygastroidea
E3F10R1	187	26	25	88	166		Chalcidoidea
E3F9R1	187	26	25	220	166		Chalcidoidea
E1F5R8	188	26	25	88	166		Chalcidoidea
E6F2R2	188	26	25	92	166		Chalcidoidea
E6F3R2	188	26	25	92	166		Chalcidoidea
E6F5R5	188	26	25	92	166		Chalcidoidea
E2F3R2	188	26	25	93	166		Chalcidoidea
E4F3R6	188	26	25	93	166		Chalcidoidea
E3F7R4	188	26	25	87	167		Chalcidoidea
E3F7R5	188	26	25	87	167		Chalcidoidea

E3F12R5	188	26	25	94	168		Chalcidoidea
E3F1R1	188	26	25	94	168		Chalcidoidea
E3F3R1	188	26	25	94	168		Chalcidoidea
E3F4R3	188	26	25	94	168		Chalcidoidea
E3F5R1	188	26	25	94	168		Chalcidoidea
E2F2R2	191	63	63	46	3	S	Ichneumonoidea
E1F8R3	192	36	35	43	13		Ichneumonoidea
E2F11R5	192	36	35	43	13		Ichneumonoidea
E2F10R5	193	80	86	180	37	S	Chalcidoidea
E4F5R7	194	96	104	44	116		Ichneumonoidea
E2F9R5	194	76	83	52	127		Ichneumonoidea
E2F5R6	195	68	71	185	135	S	Chalcidoidea
E2F4R6	196	66	69	182	141	S	Chalcidoidea
E6F11R2	199	135	148	85	31		Ichneumonoidea
E6F12R2	199	135	148	85	31		Ichneumonoidea
E6F1R3	199	135	148	85	31		Ichneumonoidea
E6F2R4	199	135	148	85	31		Ichneumonoidea
E3F6R7	199	96	104	44	116		Ichneumonoidea
E6F9R2	199	76	83	52	123		Ichneumonoidea
E6F12R3	199	77	83	166	124		Ichneumonoidea
E3F5R3	199	76	83	52	125		Ichneumonoidea
E3F2R1	199	76	83	52	126		Ichneumonoidea
E2F9R6	199	76	83	52	127		Ichneumonoidea
E3F4R1	199	76	83	52	127		Ichneumonoidea
E3F9R5	199	76	83	52	127		Ichneumonoidea
E4F12R1	199	76	83	52	127		Ichneumonoidea
E4F12R5	199	76	83	52	127		Ichneumonoidea
E6F10R2	199	76	83	52	127		Ichneumonoidea
E6F1R4	199	76	83	52	127		Ichneumonoidea

E2F8R6	200	18	17	114	152		Platygastroidea
E1F4R4	200	18	17	115	152		Platygastroidea
E3F11R1	200	18	17	116	152		Platygastroidea
E4F6R4	201	124	135	97	11		Ceraphronoidea
E2F9R1	201	41	82	19	48		Ceraphronoidea
E6F2R7	201	41	82	19	48		Ceraphronoidea
E6F3R7	201	41	82	19	48		Ceraphronoidea
E6F4R7	201	41	82	19	48		Ceraphronoidea
E6F7R4	201	41	82	19	48		Ceraphronoidea
E4F1R4	201	41	41	11	49		Ceraphronoidea
E4F5R3	201	41	41	11	49		Ceraphronoidea
E4F8R4	201	41	41	11	49		Ceraphronoidea
E1F11R6	201	41	41	105	49		Ceraphronoidea
E6F5R7	201	41	41	10	51		Ceraphronoidea
E6F6R5	201	41	41	10	51		Ceraphronoidea
E6F9R4	201	41	41	10	51		Ceraphronoidea
E1F12R1	201	41	41	14	52		Ceraphronoidea
E2F10R7	201	41	41	14	52		Ceraphronoidea
E1F9R4	201	41	41	15	52		Ceraphronoidea
E2F2R3	201	41	41	15	52		Ceraphronoidea
E6F1R5	201	41	41	10	52		Ceraphronoidea
E3F6R3	201	41	41	13	52		Ceraphronoidea
E6F4R4	201	41	41	13	52		Ceraphronoidea
E6F8R4	201	41	41	13	52		Ceraphronoidea
E4F9R6	201	41	41	13	54		Ceraphronoidea
E4F2R5	201	116	127	103	55		Ceraphronoidea
E4F2R6	201	116	127	104	55		Ceraphronoidea
E5F12R5	201	120	131	17	56		Ceraphronoidea
E5F11R5	201	41	41	16	189		Ceraphronoidea

E5F8R4	201	41	41	16	189		Ceraphronoidea
E6F9R5	201	41	41	10	190		Ceraphronoidea
E6F3R4	201	41	41	13	190		Ceraphronoidea
E6F5R4	201	41	41	13	190		Ceraphronoidea
E6F6R4	201	41	41	13	191		Ceraphronoidea
E3F7R1	202	97	105	42	60		Platygastroidea
E6F12R5	202	97	105	42	60		Platygastroidea
E3F7R3	202	98	107	41	61		Platygastroidea
E3F8R7	202	98	107	41	61		Platygastroidea
E6F6R7	202	97	105	42	62		Platygastroidea
E6F9R7	202	97	105	42	63		Platygastroidea
E3F3R7	202	40	40	127	67		Platygastroidea
E1F11R3	202	52	52	113	128		Platygastroidea
E1F9R2	202	40	40	112	147		Platygastroidea
E3F4R6	202	40	65	130	192		Platygastroidea
E3F2R6	202	40	65	36	193		Platygastroidea
E5F3R7	202	40	65	36	193		Platygastroidea
E5F4R7	202	40	65	36	193		Platygastroidea
E5F5R7	202	40	65	36	193		Platygastroidea
E5F6R7	202	40	65	36	193		Platygastroidea
E2F3R8	202	40	65	37	193		Platygastroidea
E3F11R7	202	40	65	37	193		Platygastroidea
E6F11R5	202	40	65	37	193		Platygastroidea
E6F7R7	202	40	65	37	193		Platygastroidea
E6F8R7	202	40	65	37	193		Platygastroidea
E4F7R7	202	40	65	131	193		Platygastroidea
E3F8R3	203	99	109	148	118		Cynipoidea
E4F2R2	203	115	126	40	119		Cynipoidea
E6F10R7	203	115	126	40	119		Cynipoidea

E1F8R4	204	37	36	150	42	S	Ichneumonoidea
E1F11R7	205	54	54	117	158		Platygastroidea
E1F5R4	205	22	21	30	159		Platygastroidea
E5F9R4	205	22	21	30	159		Platygastroidea
E3F9R3	206	26	25	88	46	S	Chalcidoidea
E3F10R3	207	103	114	215	32	S	Chalcidoidea
E1F7R4	209	32	31	196	163	S	Chalcidoidea
E3F11R3	210	108	119	189	184		Chalcidoidea
E4F10R6	210	108	118	188	185		Chalcidoidea
E3F12R3	211	112	123	61	142		Chalcidoidea
E5F11R7	211	112	123	61	142		Chalcidoidea
E5F2R5	212	63	63	46	3		Ichneumonoidea
E1F7R3	212	31	30	45	117		Ichneumonoidea
E4F5R4	212	31	30	45	117		Ichneumonoidea
E5F10R4	212	31	30	45	117		Ichneumonoidea
E5F11R4	212	31	30	45	117		Ichneumonoidea
E5F12R4	212	31	30	45	117		Ichneumonoidea
E5F1R5	212	31	30	45	117		Ichneumonidea
E5F3R5	212	31	30	45	117		Ichneumonoidea
E5F4R5	212	31	30	45	117		Ichneumonoidea
E5F5R5	212	31	30	45	117		Ichneumonoidea
E5F6R5	212	31	30	45	117		Ichneumonoidea
E4F8R5	212	126	137	47	165		Ichneumonoidea
E6F11R7	212	126	137	47	165		Ichneumonoidea
E6F12R7	212	126	137	47	165		Ichneumonoidea
E1F1R8	213	2	2	38	94		Platygastroidea
E1F1R4	213	2	2	140	94		Platygastroidea
E3F11R4	213	109	120	172	153		Platygastroidea
E3F1R4	213	89	95	54	154		Platygastroidea

E4F10R5	213	89	95	54	154		Platygastroidea
E3F2R4	214	91	97	60	16		Chalcidoidea
E3F8R6	214	91	97	178	16		Chalcidoidea
E4F4R6	214	26	25	91	166		Chalcidoidea
E4F7R4	214	26	25	91	166		Chalcidoidea
E4F9R5	214	26	25	91	166		Chalcidoidea
E1F6R7	214	26	25	87	167		Chalcidoidea
E2F3R4	214	26	25	87	167		Chalcidoidea
E3F7R6	215	57	57	20	1		Ceraphronoidea
E6F4R3	215	41	41	11	49		Ceraphronoidea
E4F4R2	215	120	131	17	56		Ceraphronoidea
E1F8R6	215	39	38	98	58		Ceraphronoidea
E2F1R2	215	59	59	144	74		Ceraphronoidea
E3F10R6	215	105	116	3	122		Ceraphronoidea
E4F2R3	215	105	116	3	122		Ceraphronoidea
E4F11R5	215	62	141	22	181		Ceraphronoidea
E4F12R6	215	62	141	22	183		Ceraphronoidea
E1F9R5	216	42	42	156	12	S	Ichneumonidea
E4F5R2	217	122	133	161	20	S	Ichneumonidea
E2F12R8	218	87	93	143	75	S	Ceraphronidea
E2F9R3	220	53	53	207	29		Diaprioidea
E4F6R5	220	33	32	82	72		Diaprioidea
E4F6R6	220	33	32	82	72		Diaprioidea
E1F7R5	220	33	32	212	72		Diaprioidea
E2F7R3	220	71	74	209	101		Diaprioidea
E1F10R2	220	7	6	80	102		Diaprioidea
E3F8R4	221	49	110	187	137	S	Chalcidoidea
E3F8R5	222	100	111	139	172		Ichneumonoidea
E3F2R5	222	92	98	138	173		Ichneumonoidea

E3F12R7	223	113	124	65	180		Chalcidoidea
E4F5R6	223	113	124	65	180		Chalcidoidea
E3F10R5	224	104	115	147	22	S	Platygastroidea
E4F5R5	225	23	106	191	15	S	Chalcidoidea
E3F1R8	226	90	96	159	186		Ichneumonoidea
E3F7R7	226	90	108	158	187		Ichneumonoidea
E3F3R8	228	36	35	43	13	S	Ichneumonoidea
E2F8R8	229	75	81	135	36	S	Ichneumonoidea
E2F5R3	234	67	70	176	39	S	Ichneumonoidea
E2F1R4	235	61	61	194	179	S	Chalcidoidea
E2F7R1	236	70	73	58	140		Chalcidoidea
E5F9R7	236	70	73	58	140		Chalcidoidea
E2F2R4	238	40	40	29	147	S	Platygastroidea

Bibliographic references

- Abadie, J.-C., Andrade, C., Machon, N., & Porcher, E. (2008). On the use of parataxonomy in biodiversity monitoring: a case study on wild flora. *Biodiversity and Conservation*, *17*(14), 3485–3500.
- Aguiar, A. P., Deans, A. R., Engel, M. S., Forshage, M., Huber, J. T., Jennings, J. T., Johnson, N. F., Lelej, A. S., Longino, J. T., Lohrmann, V., Mikó, I., Ohl, M., Rasmussen, C., Taeger, A., & Yu, D. S. K. (2013). Order Hymenoptera. In: Zhang, Z.-Q. (Ed.) Animal Biodiversity: An Outline of Higher-level Classification and Survey of Taxonomic Richness. *Zootaxa*, *3703*(1), 51.
- Ahrens, D., Fujisawa, T., Krammer, H.-J., Eberle, J., Fabrizi, S., & Vogler, A. P. (2016). Rarity and Incomplete Sampling in DNA-Based Species Delimitation. *Systematic Biology*, *65*(3), 478–494.
- Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*(1), 134–147.
- Amberg, J. J., Grace McCalla, S., Monroe, E., Lance, R., Baerwaldt, K., & Gaikowski, M. P. (2015). Improving efficiency and reliability of environmental DNA analysis for silver carp. *Journal of Great Lakes Research, 41(2), 367–373.*
- Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., Kightley, E. P., Thompson, L. R., Hyde, E. R., Gonzalez, A., & Knight, R. (2017). Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *MSystems*, *2*(2), e00191-16.
- Anderson, C., Low-Choy, S., Whittle, P., Taylor, S., Gambley, C., Smith, L., Gillespie, P., Löcker, H., Davis, R., & Dominiak, B. (2017). Australian plant biosecurity surveillance systems. *Crop Protection*, *100*(00), 8–20.
- Andrew, R. L., Bernatchez, L., Bonin, A., Buerkle, C. A., Carstens, B. C., Emerson, B. C., Garant, D., Giraud, T., Kane, N. C., Rogers, S. M., Slate, J., Smith, H., Sork, V. L., Stone, G. N., Vines, T. H., Waits, L., Widmer, A., & Rieseberg, L. H. (2013). A road map for molecular ecology. *Molecular Ecology*, *22*(10), 2605–2626.
- Andújar, C., Arribas, P., Yu, D. W., Vogler, A. P., & Emerson, B. C. (2018). Why the COI barcode should be the community DNA metabarcode for the metazoa. *Molecular Ecology*, *27*(20), 3968–3975.

- Ärje, J., Melvad, C., Jeppesen, M. R., Madsen, S. A., Raitoharju, J., Rasmussen, M. S., Iosifidis, A., Tirronen, V., Meissner, K., Gabbouj, M., & Høye, T. T. (2020). Automatic image-based identification and biomass estimation of invertebrates. *ArXiv:2002.03807 [Cs, Stat]*. 1-19.
- Arulandhu, A. J., Staats, M., Hagelaar, R., Voorhuijzen, M. M., Prins, T. W., Scholtens, I., Costessi, A., Duijsings, D., Rechenmann, F., Gaspar, F. B., Barreto Crespo, M. T., Holst-Jensen, A., Birck, M., Burns, M., Haynes, E., Hochegger, R., Klingl, A., Lundberg, L., Natale, C., Kok, E. (2017). Development and validation of a multi-locus DNA metabarcoding method to identify endangered species in complex samples. *GigaScience*, *6*(10).
- Asghar, U., Malik, M. F., Anwar, F., Javed, A., & Raza, A. (2015). DNA Extraction from Insects by Using Different Techniques: A Review. *Advances in Entomology*, *03*(04), 132–138.
- Ashfaq, M., & Hebert, P. D. N. (2016). DNA barcodes for bio-surveillance: regulated and economically important arthropod plant pests. *Genome*, *59*(11), 933–945.
- Aylagas, E., Borja, Á., Irigoien, X., & Rodríguez-Ezpeleta, N. (2016). Benchmarking DNA Metabarcoding for Biodiversity-Based Monitoring and Assessment. *Frontiers in Marine Science*, *3*(00), 1-12.
- Bahder, B. W., Bollinger, M. L., Sudarshana, M. R., & Zalom, F. G. (2015). Preparation of Mealybugs (Hemiptera: Pseudococcidae) for Genetic Characterization and Morphological Examination. *Journal of Insect Science*, *15*(1), 1-5.
- Bar-On Y.M., Phillips R., Milo R. 2018. The biomass distribution on Earth. Proc. *Natl. Acad. Sci. 115*, 6506–6511.
- Barrett, R. D. H., & Hebert, P. D. N. (2005). Identifying spiders through DNA barcodes. *Canadian Journal of Zoology*, *83*(3), 481–491. https://doi.org/10.1139/z05-024
- Barroso, P. C. S., Menezes, R. S. T., de Oliveira, M. L., & Somavilla, A. (2022). A systematic review of the Neotropical social wasp genus Angiopolybia Araujo, 1946 (Hymenoptera: Vespidae): species delimitation, morphological diagnosis, and geographical distribution. *Arthropod Systematics & Phylogeny*, *80*, 75–97.
- Barwell, L. J., Isaac, N. J. B., & Kunin, W. E. (2015). Measuring β - diversity with species abundance data. *Journal of Animal Ecology*, *84*(4), 1112–1122.

- Batovska, J., Lynch, S. E., Cogan, N. O. I., Brown, K., Darbro, J. M., Kho, E. A., & Blacket, M. J. (2018). Effective mosquito and arbovirus surveillance using metabarcoding. *Molecular Ecology Resources*, *18*(1), 32–40.
- Batovska, J., Piper, A. M., Valenzuela, I., Cunningham, J. P., & Blacket, M. J. (2021). Developing a non-destructive metabarcoding protocol for detection of pest insects in bulk trap catches. *Scientific Reports*, *11*(1), 7946.
- Behrens-Chapuis, S., Herder, F., & Geiger, M. F. (2021). Adding DNA barcoding to stream monitoring protocols – What's the additional value and congruence between morphological and molecular identification approaches? *PLOS ONE*, *16*(1), e0244598.
- Belokobyl'skij, S. A., & Lelej, A. S. (2019). *Annotated catalogue of the Hymenoptera of Russia. Volume 2 Volume 2*. Available at : https://www.zin.ru/journals/trudyzin/doc/vol_323_s8/TZ_323_8_Supplement_Belokobylskij-Lelej.pdf
- Belokobylskij, S. A., Samartsev, K. G., & Il'inskaya, A. S. (Eds.). (2019). Annotated catalogue of the Hymenoptera of Russia. Volume II. Apocrita: Parasitica. *Proceedings of the Zoological Institute RAS*, *323*(Supplement 8), 1–594.
- Bennett, A. M. R., Cardinal, S., Gauld, I. D., & Wahl, D. B. (2019) b. Phylogeny of the subfamilies of Ichneumonidae (Hymenoptera). *Journal of Hymenoptera Research*, *71*, 1–156.
- Bennett, A. M. R., Sheffield, C. S., & deWaard, J. R. (2019) a. Hymenoptera of Canada. *ZooKeys*, *819*, 311–360.
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, *456*(7218), 53–59.
- Bik, H. M., Porazinska, D. L., Creer, S., Caporaso, J. G., Knight, R., & Thomas, W. K. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution*, *27*(4), 233–243.
- Binladen, J., Thomas, M., Gilbert, P., Campos, P. F., & Willerslev, E. (2007). 5′-Tailed sequencing primers improve sequencing quality of PCR products. *BioTechniques*, *42*(2), 174–176.

- Bohmann, K., Gopalakrishnan, S., Nielsen, M., Nielsen, L. dos S. B., Jones, G., Streicker, D. G., & Gilbert, M. T. P. (2018). Using DNA metabarcoding for simultaneous inference of common vampire bat diet and population structure. *Molecular Ecology Resources*, *18*(5), 1050–1063.
- Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., Mills, D. A., & Caporaso, J. G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, *10*(1), 57–59.
- Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating pearson, kendall and spearman correlations. *Psychometrika*, *65*(1), 23–28.
- Borisenko, A. V., Sones, J. E., & Hebert, P. D. N. (2009). The front-end logistics of DNA barcoding: challenges and prospects. *Molecular Ecology Resources*, *9*(1), 27–34.
- Boslaugh, Sarah and Paul Andrew Watters. 2008. *Statistics in a Nutshell: A Desktop Quick Reference*, ch. 16. Sebastopol, CA: O'Reilly Media. ISBN-13: 978.
- Brandon-Mong, G.-J., Gan, H.-M., Sing, K.-W., Lee, P.-S., Lim, P.-E., & Wilson, J.-J. (2015). DNA metabarcoding of insects and allies: an evaluation of primers and pipelines. *Bulletin of Entomological Research*, *105*(6), 717–727.
- Branstetter, M. G., Danforth, B. N., Pitts, J. P., Faircloth, B. C., Ward, P. S., Buffington, M. L., Gates, M. W., Kula, R. R., & Brady, S. G. (2017). Phylogenomic Insights into the Evolution of Stinging Wasps and the Origins of Ants and Bees. *Current Biology*, *27*(7).
- Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J., Elbrecht, V., Steinke, D., Ratnasingham, S., Waard, J. R. de, Sones, J. E., Zakharov, E. V., & Hebert, P. D. N. (2019). Metabarcoding a diverse arthropod mock community. *Molecular Ecology Resources*, *19*(3), 711–727.
- Buée, M., Reich, M., Murat, C., Morin, E., Nilsson, R. H., Uroz, S., & Martin, F. (2009). 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity: Research. *New Phytologist*, *184*(2), 449–456.
- Büsse, S., von Grumbkow, P., Mazanec, J., Tröster, G., Hummel, S., & Hörnschemeyer, T. (2017). Note on using nuclear 28S rDNA for sequencing ancient and strongly degraded insect DNA. *Entomological Science, 20*(1), 137–141.
- Calderón-Cortés, N., Quesada, M., Cano-Camacho, H., & Zavala-Páramo, G. (2010). A Simple and Rapid Method for DNA Isolation from Xylophagous Insects. *International Journal of Molecular Sciences*, *11*(12), 5056–5064.

- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7).
- Callahan, B. J., Wong, J., Heiner, C., Oh, S., Theriot, C. M., Gulati, A. S., McGill, S. K., & Dougherty, M. K. (2019). High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Research*, *47*(18), 1-12.
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*(5), 335–336.
- Carbayo, F., & Marques, A. C. (2011). The costs of describing the entire animal kingdom. *Trends in Ecology & Evolution*, *26*(4), 154–155.
- Carew, M. E., Coleman, R. A., & Hoffmann, A. A. (2018). Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding? *PeerJ*, *6*, e4980.
- Carew, M. E., Nichols, S. J., Batovska, J., St Clair, R., Murphy, N. P., Blacket, M. J., & Shackleton, M. E. (2017). A DNA barcode database of Australia's freshwater macroinvertebrate fauna. *Marine and Freshwater Research*, *68*(10), 1788-1802.
- Casetta, E., Marques da Silva, J., & Vecchi, D. (Eds.). (2019). *From Assessing to Conserving Biodiversity: Conceptual and Practical Challenges* (Vol. 24). Springer International Publishing.
- Casquet, J., Thebaud, C., & Gillespie, R. G. (2012). Chelex without boiling, a rapid and easy technique to obtain stable amplifiable DNA from small amounts of ethanol-stored spiders: QUICK AND EASY DSDNA EXTRACTION FOR BARCODING. *Molecular Ecology Resources*, *12*(1), 136–141.
- Castalanelli, M. A., Severtson, D. L., Brumley, C. J., Szito, A., Foottit, R. G., Grimm, M., Munyard, K., & Groth, D. M. (2010). A rapid non-destructive DNA extraction method for insects and other arthropods. *Journal of Asia-Pacific Entomology*, *13*(3), 243–248.
- Caterino, M. S., Cho, S., & Sperling, F. A. H. (2000). The Current State Of Insect Molecular Systematics: A Thriving Tower of Babel. *Annual Review of Entomology*, *45*(1), 1–54.
- Chang, J. J. M., Ip, Y. C. A., Bauman, A. G., & Huang, D. (2020). MinION-in-ARMS: Nanopore Sequencing to Expedite Barcoding of Specimen-Rich Macrofaunal Samples From Autonomous Reef Monitoring Structures. *Frontiers in Marine Science*, *7*(448), 1-13.

- Chang, J. J. M., Ip, Y. C. A., Cheng, L., Kunning, I., Mana, R. R., Wainwright, B. J., & Huang, D. (2022). High-Throughput Sequencing for Life-History Sorting and for Bridging Reference Sequences in Marine Gerromorpha (Insecta: Heteroptera). *Insect Systematics and Diversity*, *6*(1), 1.
- Chapman, A. (2009). Numbers of living species in Australia and the world, 2nd edn. Department of Sustainability, Environment, Water, Population and Communities, Canberra, Australia.
- Chen, H., Rangasamy, M., Tan, S. Y., Wang, H., & Siegfried, B. D. (2010). Evaluation of Five Methods for Total DNA Extraction from Western Corn Rootworm Beetles. *PLoS ONE*, *5*(8), 1-6.
- Chen, M., Zhu, Y., Tao, J., & Luo, Y. (2008). Methodological comparison of DNA extraction from Holcocerrus hippophaecolus (Lepidoptera: Cossidae) for AFLP analysis. *Forestry Studies in China*, *10*(3), 189–192.
- Chen, R., Jiang, L.-Y., & Qiao, G.-X. (2012). The Effectiveness of Three Regions in Mitochondrial Genome for Aphid DNA Barcoding: A Case in Lachininae. *PLoS ONE*, *7*(10), 1-11.
- Chen, R., Jiang, L.-Y., Liu, L., Liu, Q.-H., Wen, J., Zhang, R.-L., Li, X.-Y., Wang, Y., Lei, F.-M., & Qiao, G.-X. (2013). The gnd gene of Buchnera as a new, effective DNA barcode for aphid identification. *Systematic Entomology*, *38*(3), 615–625.
- Chiu, C. Y., & Miller, S. A. (2019). Clinical metagenomics. *Nature Reviews. Genetics*, *20*(6), 341–355.
- Cho, S., Mitchell, A., Regier, J. C., Mitter, C., Poole, R. W., Friedlander, T. P., & Zhao, S. (1995). A highly conserved nuclear gene for low-level phylogenetics: elongation factor-1 alpha recovers morphology-based tree for heliothine moths. *Molecular Biology and Evolution*, *12*(4), 650–656.
- Choi, H., Shin, S., Jung, S., Clarke, D. J., & Lee, S. (2018). Molecular phylogeny of Macrosiphini (Hemiptera: Aphididae): An evolutionary hypothesis for the Pterocomma-group habitat adaptation. *Molecular Phylogenetics and Evolution*, *121*(00), 12–22.
- Clarke, L. J., Beard, J. M., Swadling, K. M., & Deagle, B. E. (2017). Effect of marker choice and thermal cycling protocol on zooplankton DNA metabarcoding studies. *Ecology and Evolution*, *7*(3), 873–883.

- Clarke, L. J., Soubrier, J., Weyrich, L. S., & Cooper, A. (2014). Environmental metabarcodes for insects: *in silico* PCR reveals potential for taxonomic bias. *Molecular Ecology Resources*, *14*(6), 1160–1170.
- Condamine, F. L., Toussaint, E. F. A., Cotton, A. M., Genson, G. S., Sperling, F. A. H., & Kergoat, G. J. (2013). Fine-scale biogeographical and temporal diversification processes of peacock swallowtails (Papilio subgenus Achillides) in the Indo-Australian Archipelago. *Cladistics*, *29*(1), 88–111.
- Cooper, S. J. B., Watts, C. H. S., Saint, K. M., & Leijs, R. (2014). Phylogenetic relationships of Australian Scirtidae (Coleoptera) based on mitochondrial and nuclear sequences. *Invertebrate Systematics*, *28*(6), 628–642.
- Cornils, A. (2015). Non-destructive DNA extraction for small pelagic copepods to perform integrative taxonomy. *Journal of Plankton Research*, *37*(1), 6–10.
- Costello, M., Fleharty, M., Abreu, J., Farjoun, Y., Ferriera, S., Holmes, L., Granger, B., Green, L., Howd, T., Mason, T., Vicente, G., Dasilva, M., Brodeur, W., DeSmet, T., Dodge, S., Lennon, N. J., & Gabriel, S. (2018). Characterization and remediation of sample index swaps by non-redundant dual indexing on massively parallel sequencing platforms. *BMC Genomics*, *19*(1), 1-10.
- Crampton-Platt, A., Yu, D. W., Zhou, X., & Vogler, A. P. (2016). Mitochondrial metagenomics: letting the genes out of the bottle. *GigaScience*, *5*(1), 1-15.
- Creedy, T. J., Norman, H., Tang, C. Q., Chin, K. Q., Andujar, C., Arribas, P., O'Connor, R. S., Carvell, C., Notton, D. G., & Vogler, A. P. (2020). A validated workflow for rapid taxonomic assignment and monitoring of a national fauna of bees (Apiformes) using high throughput DNA barcoding. *Molecular Ecology Resources*, *20*(1), 40–53.
- Creer, S., Fonseca, V. G., Porazinska, D. L., Giblin-Davis, R. M., Sung, W., Power, D. M., Packer, M., Carvalho, G. R., Blaxter, M. L., Lambshead, P. J. D., & Thomas, W. K. (2010). Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology*, *19*(1), 4–20.
- Cruaud, A., Groussier, G., Genson, G., Sauné, L., Polaszek, A., & Rasplus, J.-Y. (2018). Pushing the limits of whole genome amplification: successful sequencing of RADseq library from a single microhymenopteran (Chalcidoidea, *Trichogramma*). *PeerJ*, *6*, e5640.
- Cruaud, A., Nidelet, S., Arnal, P., Weber, A., Fusu, L., Gumovsky, A., Huber, J., Polaszek, A., & Rasplus, J. (2019). Optimized DNA extraction and library preparation for minute

arthropods: Application to target enrichment in chalcid wasps used for biocontrol. *Molecular Ecology Resources*, *19*(3), 702–710.

- Cruaud, P., Rasplus, J.-Y., Rodriguez, L. J., & Cruaud, A. (2017). High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. *Scientific Reports*, *7*(1), 1-12.
- Cywinska, A., Hunter, F. F., & Hebert, P. D. N. (2006). Identifying Canadian mosquito species through DNA barcodes. *Medical and Veterinary Entomology*, *20*(4), 413–424.
- D'acier, A. C., Cruaud, A., Artige, E., Genson, G., Clamens, A.-L., Pierre, E., Hudaverdian, S., Simon, J.-C., Jousselin, E., & Rasplus, J.-Y. (2014). DNA Barcoding and the Associated PhylAphidB@se Website for the Identification of European Aphids (Insecta: Hemiptera: Aphididae). *PLOS ONE*, *9*(6), 1-16.
- de Kerdrel, G. A., Andersen, J. C., Kennedy, S. R., Gillespie, R., & Krehenwinkel, H. (2020). Rapid and cost-effective generation of single specimen multilocus barcoding data from whole arthropod communities by multiple levels of multiplexing. *Scientific Reports*, *10*(1), 1-12.
- de la Cruz-Ramos, J. M., Hernández-Triana, L. M., García-De la Peña, C., González-Álvarez, V. H., Weger-Lucarelli, J., Siller-Rodríguez, Q. K., Sánchez Rámos, F. J., Rodríguez, A. D., & Ortega-Morales, A. I. (2019). Comparison of two DNA extraction methods from larvae, pupae, and adults of Aedes aegypti. *Heliyon*, *5*(10), e02660.
- Deagle, B. E., Eveson, J. P., & Jarman, S. N. (2006). Quantification of damage in DNA recovered from highly degraded samples – a case study on DNA in faeces. *Frontiers in Zoology*, *3*(1), 1-10.
- Deagle, B. E., Thomas, A. C., McInnes, J. C., Clarke, L. J., Vesterinen, E. J., Clare, E. L., Kartzinel, T. R., & Eveson, J. P. (2019). Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular Ecology*, *28*(2), 391–406.
- Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., Vere, N. de, Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895.
- Dentinger, B. T. M., Margaritescu, S., & Moncalvo, J.-M. (2009). Rapid and reliable high-throughput methods of DNA extraction for use in barcoding and molecular systematics of mushrooms: DNA BARCODING. *Molecular Ecology Resources*, *10*(4), 628–633.

- Depa, Ł., Kanturski, M., Taszakowski, A., Walczak, M., Taszakowski, A., Walczak, M., Bugaj-Nawrocka, A., & Wieczorek, K. (2017). Dysaphis (Dysaphis) kadyrovi sp. nov.—a new aphid species (Hemiptera: Aphididae) from Tajikistan. *Zootaxa*, *4286*(4), 573–585.
- Derraik, J. G. B., Barratt, B. I. P., Sirvid, P., Macfarlane, R. P., Patrick, B. H., Early, J., Eyles, A. C., Johns, P. M., Fraser, P. M., Barker, G. M., Henderson, R., Dale, P. J., Harvey, M. S., Fenwick, G., Mclellan, I. D., Dickinson, K. J. M., & Closs, G. P. (2001). Invertebrate survey of a modified native shrubland, Brookdale Covenant, Rock and Pillar Range, Otago, New Zealand. *New Zealand Journal of Zoology*, *28*(3), 273–290.
- Derraik, J. G. B., Closs, G. P., Dickinson, K. J. M., Sirvid, P., Barratt, B. I. P., & Patrick, B. H. (2002). Arthropod Morphospecies versus Taxonomic Species: a Case Study with Araneae, Coleoptera, and Lepidoptera. *Conservation Biology*, *16*(4), 1015–1023.
- Derraik, J. G. B., Early, J. W., Closs, G. P., & Dickinson, K. J. M. (2010). Morphospecies and Taxonomic Species Comparison for Hymenoptera. *Journal of Insect Science*, *10*(108), 1–7.
- Diekstra, A., Bosgoed, E., Rikken, A., van Lier, B., Kamsteeg, E.-J., Tychon, M., Derks, R. C., van Soest, R. A., Mensenkamp, A. R., Scheffer, H., Neveling, K., & Nelen, M. R. (2015). Translating Sanger-Based Routine DNA Diagnostics into Generic Massive Parallel Ion Semiconductor Sequencing. *Clinical Chemistry*, *61*(1), 154–162.
- Dimitrov, D., Benavides, L. R., Arnedo, M. A., Giribet, G., Griswold, C. E., Scharff, N., & Hormiga, G. (2017). Rounding up the usual suspects: a standard target-gene approach for resolving the interfamilial phylogenetic relationships of ecribellate orb-weaving spiders with a new family-rank classification (Araneae, Araneoidea). *Cladistics*, *33*(3), 221–250.
- Dincă, V., Montagud, S., Talavera, G., Hernández-Roldán, J., Munguira, M. L., García-Barros, E., Hebert, P. D. N., & Vila, R. (2015). DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. *Scientific Reports*, *5*(1), 12395. https://doi.org/10.1038/srep12395
- Djurhuus, A., Port, J., Closek, C. J., Yamahara, K. M., Romero-Maraccini, O., Walz, K. R., Goldsmith, D. B., Michisaki, R., Breitbart, M., Boehm, A. B., & Chavez, F. P. (2017). Evaluation of Filtration and DNA Extraction Methods for Environmental DNA Biodiversity Assessments across Multiple Trophic Levels. *Frontiers in Marine Science*, *4*(314), 1-11.
- Dohm, J. C., Lottaz, C., Borodina, T., & Himmelbauer, H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*, *36*(16), 1-10.

- Dopheide, A., Tooman, L. K., Grosser, S., Agabiti, B., Rhode, B., Xie, D., Stevens, M. I., Nelson, N., Buckley, T. R., Drummond, A. J., & Newcomb, R. D. (2019). Estimating the biodiversity of terrestrial invertebrates on a forested island using DNA barcodes and metabarcoding data. *Ecological Applications*, *29*(4), 1-14.
- Dreo, T., Pirc, M., Ramšak, Ž., Pavšič, J., Milavec, M., Žel, J., & Gruden, K. (2014). Optimising droplet digital PCR analysis approaches for detection and quantification of bacteria: a case study of fire blight and potato brown rot. *Analytical and Bioanalytical Chemistry*, *406*(26), 6513–6528.
- Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed Phylogenetics and Dating with Confidence. *PLoS Biology*, *4*(5), e88.
- Ebach, M. C., & Holdrege, C. (2005). DNA barcoding is no substitute for taxonomy. *Nature*, *434*(7034), 697–697.
- Eberle, J., Ahrens, D., Mayer, C., Niehuis, O., & Misof, B. (2020). A Plea for Standardized Nuclear Markers in Metazoan DNA Taxonomy. *Trends in Ecology & Evolution*, *35*(4), 336–345.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340
- Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*(10), 996–998. https://doi.org/10.1038/nmeth.2604
- Edgar, R. C. (2016). SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. *BioRxiv*, *7*(00),1-20.
- Edgar, R. C. (2016). UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *Bioinformatics, 9*(101), 1-21.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., deWinter, A., Dixon, J., Turner, S. (2009). Real-Time DNA Sequencing from Single Polymerase Molecules. *Science*, *323*(5910), 133–138.
- Elbrecht, V., & Leese, F. (2015). Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass—Sequence Relationships with an Innovative Metabarcoding Protocol. *PLOS ONE*, *10*(7), 1-16.
- Elbrecht, V., & Leese, F. (2017). PrimerMiner: an r package for development and in silico validation of DNA metabarcoding primers. *Methods in Ecology and Evolution*, *8*(5), 622–626.

- Elbrecht, V., Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J., Hajibabaei, M., Wright, M., Zakharov, E. V., Hebert, P. D. N., & Steinke, D. (2019). Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ*, *7*(00), 1-23.
- Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., & Leese, F. (2017). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, *8*(10), 1265–1275.
- Evans, J. D., Schwarz, R. S., Chen, Y. P., Budge, G., Cornman, R. S., De la Rua, P., de Miranda, J. R., Foret, S., Foster, L., Gauthier, L., Genersch, E., Gisder, S., Jarosch, A., Kucharski, R., Lopez, D., Lun, C. M., Moritz, R. F. A., Maleszka, R., Muñoz, I., & Pinto, M. A. (2013). Standard methods for molecular research in *Apis mellifera*. *Journal of Apicultural Research*, *52*(4), 1–54.
- Fagan-Jeffries, E. P., Cooper, S. J. B., Bertozzi, T., Bradford, T. M., & Austin, A. D. (2018). DNA barcoding of microgastrine parasitoid wasps (Hymenoptera: Braconidae) using high-throughput methods more than doubles the number of species known for Australia. *Molecular Ecology Resources*, *18*(5), 1132–1143.
- Fagan-Jeffries, E. P., Cooper, S. J. B., Bradford, T. M., & Austin, A. D. (2019). Intragenomic *internal transcribed spacer 2* variation in a genus of parasitoid wasps (Hymenoptera: Braconidae): implications for accurate species delimitation and phylogenetic analysis. *Insect Molecular Biology*, imb.12564.
- Fan, J. A., Gu, H., Chen, S., Mo, B., Wen, Y., He, W., & Zeng, X. (2009). Species 637 identification of 36 kinds of fruit flies based on minimalist-barcode. *Chinese Journal of Applied 638 & Environmental Biology*, *2* (00), 215-219.
- Favret, C. (2005) A new non-destructive DNA extraction and specimen clearing technique for aphids (Hemiptera). *Proceedings of the Entomological Society of* Washington *107*(2), 469–470
- Favret, C., & Voegtlin, D. J. (2004). Speciation by host-switching in pinyon Cinara (Insecta: Hemiptera: Aphididae). *Molecular Phylogenetics and Evolution*, *32*(1), 139–151.
- Favret, C., Lessard, V., Trépanier, A., Eon-Le Guern, T., & Théry, T. (2019). Voegtlin-style suction traps measure insect diversity and community heterogeneity. *Insect Conservation and Diversity*, *12*(5), 373–381.
- Fernández-Flores, S., Fernández-Triana, J. L., Martínez, J. J., & Zaldívar-Riverón, A. (2013). DNA barcoding species inventory of Microgastrinae wasps (Hymenoptera,

Braconidae) from a Mexican tropical dry forest. *Molecular Ecology Resources*, *42*(5), 1-18.

- Fields, A. T., Abercrombie, D. L., Eng, R., Feldheim, K., & Chapman, D. D. (2015). A Novel Mini-DNA Barcoding Assay to Identify Processed Fins from Internationally Protected Shark Species. *PLOS ONE*, *10*(2), 1-10.
- Floyd, R., Lima, J., deWaard, J., Humble, L., & Hanner, R. (2010). Common goals: policy implications of DNA barcoding as a protocol for identification of arthropod pests. *Biological Invasions*, *12*(9), 2947–2954.
- Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299.
- Fontaneto, D., Flot, J.-F., & Tang, C. Q. (2015). Guidelines for DNA taxonomy, with a focus on the meiofauna. *Marine Biodiversity*, *45*(3), 433–451.
- Forbes, A. A., Bagley, R. K., Beer, M. A., Hippee, A. C., & Widmayer, H. A. (2018). Quantifying the unquantifiable: why Hymenoptera, not Coleoptera, is the most speciose animal order. *BMC Ecology*, *18*(1), 21.
- Freeland, J. R. (2017). The importance of molecular markers and primer design when characterizing biodiversity from environmental DNA. *Genome, 60*(4), 358–374.
- Fujisawa, T., & Barraclough, T. G. (2013). Delimiting Species Using Single-Locus Data and the Generalized Mixed Yule Coalescent Approach: A Revised Method and Evaluation on Simulated Data Sets. *Systematic Biology*, *62*(5), 707–724.
- Fukumoto, S., Ushimaru, A., & Minamoto, T. (2015). A basin-scale application of environmental DNA assessment for rare endemic species and closely related exotic species in rivers: a case study of giant salamanders in Japan. *Journal of Applied Ecology*, *52*(2), 358–365.
- Gauld, I. D., & Fitton, M. G. (1987). Sexual dimorphism in Ichneumonidae: a response to Hurlbutt. *Biological Journal of the Linnean Society*, *31*(3), 291–300. https://doi.org/10.1111/j.1095-8312.1987.tb01994.x
- Geller, J., Meyer, C., Parker, M., & Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome *c* oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, *13*(5), 851–861.

- Giantsis, I. A., Chaskopoulou, A., & Bon, M. C. (2015). Mild-Vectolysis: A Non destructive DNA Extraction Method for Vouchering Sand Flies and Mosquitoes. *Journal of Medical Entomology, 53(3)*, 692–695.
- Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., Hallwachs, W., & Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasystematics. *Proceedings of the National Academy of Sciences*, *111*(22), 8007–8012.
- Gilbert, M. T. P., Moore, W., Melchior, L., & Worobey, M. (2007). DNA Extraction from Dry Museum Beetles without Conferring External Morphological Damage. *PLOS ONE*, *2*(3), 1-4.
- Giordano, F., Aigrain, L., Quail, M. A., Coupland, P., Bonfield, J. K., Davies, R. M., Tischler, G., Jackson, D. K., Keane, T. M., Li, J., Yue, J.-X., Liti, G., Durbin, R., & Ning, Z. (2017). De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. *Scientific Reports*, *7*(1), 1-10.
- Gnerre, S., Lander, E. S., Lindblad-Toh, K., & Jaffe, D. B. (2009). Assisted assembly: how to improve a de novo genome assembly by using related species. *Genome Biology*, *10*(8), R88.
- Godfray, H.C.J. (2002). Challenges for taxonomy. *Nature*, *417*(2), 17–19.
- Gohl, D. M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., Gould, T. J., Clayton, J. B., Johnson, T. J., Hunter, R., Knights, D., & Beckman, K. B. (2016). Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nature Biotechnology*, *34*(9), 942–949.
- Gompert, Z., Lucas, L. K., Buerkle, C. A., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2014). Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*, *23*(18), 4555–4573.
- Goulding, T. C., Dayrat, B. (2016). Integrative taxonomy: Ten years of practice and looking into the future. *Archives of Zoological Museum of Lomonosov Moscow State University, 54*(8), 116–33.
- Goulet, H. (1987) Symphyta. In: The insects, spiders and mites of Cape Breton Highlands National Park. *Agriculture Canada Biosystematics Research Centre Report 1, Ottawa, Ontario*, 225–230.

- Green, M. R., & Sambrook, J. (2018). Touchdown Polymerase Chain Reaction (PCR). *Cold Spring Harbor Protocols*, *2018*(5), 1-20.
- Grevelding, C. (1996). Direct PCR on fruitflies and blood flukes without prior DNA isolation. *Nucleic Acids Research*, *24*(20), 4100–4101.
- Grosdidier, M., Aguayo, J., Marçais, B., & Ioos, R. (2017). Detection of plant pathogens using real-time PCR: how reliable are late Ct values? *Plant Pathology*, *66*(3), 359–367.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321.
- Guindon, S., Lethiec, F., Duroux, P., & Gascuel, O. (2005). PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research*, *33*(6), 557–559.
- Guo, J., Xu, N., Li, Z., Zhang, S., Wu, J., Kim, D. H., Sano Marma, M., Meng, Q., Cao, H., Li, X., Shi, S., Yu, L., Kalachikov, S., Russo, J. J., Turro, N. J., & Ju, J. (2008). Four-color DNA sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(27), 9145–9150.
- Gupta, S., Preet, S. (2012). Protocol optimization for genomic DNA extraction and RAPD PCR in mosquito larvae (Diptera: Culicidae). *Ann Biol Res 3*(3):1553–1561.
- Guzmán-Larralde, A. J., Suaste-Dzul, A. P., Gallou, A., & Peña-Carrillo, K. I. (2017). DNA recovery from microhymenoptera using six non-destructive methodologies with considerations for subsequent preparation of museum slides. *Genome*, *60*(1), 85–91.
- Hajibabaei, M., Porter, T. M., Robinson, C. V., Baird, D. J., Shokralla, S., & Wright, M. (2019). Watered-down biodiversity? A comparison of metabarcoding results from DNA extracted from matched water and bulk tissue biomonitoring samples. *Ecology22*(00), 1-36.
- Hajibabaei, M., Porter, T. M., Wright, M., & Rudar, J. (2019). COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. *PLOS ONE*, *14*(9), e0220953.
- Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A. C., & Baird, D. J. (2011). Environmental Barcoding: A Next-Generation Sequencing Approach for Biomonitoring Applications Using River Benthos. *PLOS ONE*, *6*(4), 1-7.

- Hajibabaei, M., Smith, M. A., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B., & Hebert, P. D. N. (2006). A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, *6*(4), 959–964.
- Hajibabaei, M., Smith, M. A., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B., & Hebert, P. D. N. (2006). A minimalist barcode can identify a specimen whose DNA is degraded: BARCODING. *Molecular Ecology Notes*, *6*(4), 959–964.
- Hao, M., Jin, Q., Meng, G., Yang, C., Yang, S., Shi, Z., Tang, M., Liu, S., Li, Y., Li, J., Zhang, D., Su, X., Shih, C., Sun, Y., Wilson, J.-J., Zhou, X., & Zhang, A. (2020). Using full-length metabarcoding and DNA barcoding to infer community assembly for speciose taxonomic groups: a case study. *Evolutionary Ecology*, *34*(6), 1063–1088.
- Hartop, E., Häggqvist, S., Ulefors, S. O., & Ronquist, F. (2020). Scuttling towards monophyly: phylogeny of the mega-diverse genus *Megaselia* (Diptera: Phoridae). *Systematic Entomology (00)*, 1-12.
- Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques*, *56*(2), 61–77.
- Hebert, P. D. N., Braukmann, T. W. A., Prosser, S. W. J., Ratnasingham, S., deWaard, J. R., Ivanova, N. V., Janzen, D. H., Hallwachs, W., Naik, S., Sones, J. E., & Zakharov, E. V. (2018). A Sequel to Sanger: amplicon sequencing that scales. *BMC Genomics*, *19*(1), 219. 1-14.
- Hebert, P. D. N., deWaard, J. R., Zakharov, E. V., Prosser, S. W. J., Sones, J. E., McKeown, J. T. A., Mantle, B., & Salle, J. L. (2013). A DNA 'Barcode Blitz': Rapid Digitization and Sequencing of a Natural History Collection. *PLOS ONE*, *8*(7), 1-14.
- Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences*, *101*(41), 14812–14817.
- Hebert, P. D. N., Ratnasingham, S., & de Waard, J. R. (2003). Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(2), 96–99.
- Hebert, P. D. N., Ratnasingham, S., Zakharov, E. V., Telfer, A. C., Levesque-Beaudin, V., Milton, M. A., Pedersen, S., Jannetta, P., & deWaard, J. R. (2016). Counting animal species

with DNA barcodes: Canadian insects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1702), 1-10.

- Heeger, F., Bourne, E. C., Baschien, C., Yurkov, A., Bunk, B., Spröer, C., Overmann, J., Mazzoni, C. J., & Monaghan, M. T. (2018). Long-read DNA metabarcoding of ribosomal RNA in the analysis of fungi from aquatic environments. *Molecular Ecology Resources*, *18*(6), 1500–1514.
- Hernández-Triana, L. M., Crainey, J. L., Hall, A., Fatih, F., Mackenzie-Dodds, J., Shelley, A. J., Zhou, X., Post, R. J., Gregory, T. R., & Hebert, P. D. N. (2012). DNA barcodes reveal cryptic genetic diversity within the blackfly subgenus Trichodagmia Enderlein (Diptera: Simuliidae: Simulium) and related taxa in the New World. *Zootaxa*, *3514*(1), 1-43.
- Hickerson, M. J., Meyer, C. P., & Moritz, C. (2006). DNA Barcoding Will Often Fail to Discover New Animal Species over Broad Parameter Space. *Systematic Biology*, *55*(5), 729–739.
- Hirai, J., Kuriyama, M., Ichikawa, T., Hidaka, K., & Tsuda, A. (2015). A metagenetic approach for revealing community structure of marine planktonic copepods. *Molecular Ecology Resources*, *15*(1), 68–80.
- Hlaing, T., Tun-Lin, W., Somboon, P., Socheat, D., Setha, T., Min, S., Chang, M., & Walton, C. (2009). Mitochondrial pseudogenes in the nuclear genome of Aedes aegypti mosquitoes: implications for past and future population genetic studies. *BMC Genetics*, *10*(1), 1-11.
- Hoff-Olsen, P., Mevåg, B., Staalstrøm, E., Hovde, B., Egeland, T., & Olaisen, B. (1999). Extraction of DNA from decomposed human tissue. *Forensic Science International*, *105*(3), 171–183.
- Hofreiter, M., Paijmans, J. L. A., Goodchild, H., Speller, C. F., Barlow, A., Fortes, G. G., Thomas, J. A., Ludwig, A., & Collins, M. J. (2015). The future of ancient DNA: Technical advances and conceptual shifts: Prospects & Overviews. *BioEssays*, *37*(3), 284–293.
- Hong, S., Bunge, J., Leslin, C., Jeon, S., & Epstein, S. S. (2009). Polymerase chain reaction primers miss half of rRNA microbial diversity. *The ISME Journal*, *3*(12), 1365–1373.
- Hong, Y., Guo, M., & Wang, J. (2021). ENJ algorithm can construct triple phylogenetic trees. *Molecular Therapy - Nucleic Acids*, *23*, 286–293.

- Hoyt, P. R., Doktycz, M. J., Warmack, R. J., & Allison, D. P. (2001). Spin-column isolation of DNA–protein interactions from complex protein mixtures for AFM imaging. *Ultramicroscopy*, *86*(1–2), 139–143.
- Hsieh, C., Huang, C., Wu, W., & Wang, H. (2020). A rapid insect species identification system using mini-barcode pyrosequencing. *Pest Management Science*, *76*(4), 1222–1227.
- Huber, J. A., Morrison, H. G., Huse, S. M., Neal, P. R., Sogin, M. L., & Mark Welch, D. B. (2009). Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. *Environmental Microbiology*, *11*(5), 1292–1302.
- Huber, J. T. (2017). Biodiversity of Hymenoptera. In R. G. Foottit & P. H. Adler (Eds.), *Insect Biodiversity* (pp. 419–461). John Wiley & Sons, Ltd.
- Huemer, P., Mutanen, M., Sefc, K. M., & Hebert, P. D. N. (2014). Testing DNA Barcode Performance in 1000 Species of European Lepidoptera: Large Geographic Distances Have Small Genetic Impacts. *PLoS ONE*, *9*(12), e115774.
- Hunter, S. J., Goodall, T. I., Walsh, K. A., Owen, R., & Day, J. C. (2008). Non destructive DNA extraction from blackflies (Diptera: Simuliidae): retaining voucher specimens for DNA barcoding projects. *Molecular Ecology Resources*, *8*(1), 56–61.
- Hyde, K. (2013). Incorporating molecular data in fungal systematics: a guide for aspiring researchers. *Current Research in Environmental & Applied Mycology*, *3*(1), 1–32.
- Ishii, K., & Fukui, M. (2001). Optimization of Annealing Temperature To Reduce Bias Caused by a Primer Mismatch in Multitemplate PCR. *Applied and Environmental Microbiology*, *67*(8), 3753–3755.
- Ivanova, N. V., Borisenko, A. V., & Hebert, P. D. N. (2009). Express barcodes: racing from specimen to identification. *Molecular Ecology Resources*, *9*(1), 35–41.
- Ivanova, N. V., Dewaard, J. R., & Hebert, P. D. N. (2006). An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, *6*(4), 998–1002.
- Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., Tyson, J. R., Beggs, A. D., Dilthey, A. T., Fiddes, I. T., Malla, S., Marriott, H., Nieto, T., O'Grady, J., Olsen, H. E., Pedersen, B. S., Rhie, A., Richardson, H., Quinlan, A. R., Loose, M. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnology*, *36*(4), 338–345.

- Janzen, D. H., Hajibabaei, M., Burns, J. M., Hallwachs, W., Remigio, E., & Hebert, P. D. N. (2005). Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1835–1845.
- Janzen, D. H., Hallwachs, W., Pereira, G., Blanco, R., Masis, A., Chavarria, M. M., Chavarria, F., Guadamuz, A., Araya, M., Smith, M. A., Valerio, J., Guido, H., Sanchez, E., Bermudez, S., Perez, K., Manjunath, R., Ratnasingham, S., St. Jacques, B., Milton, M., Hasegawa, M. (2020). Using DNA-barcoded Malaise trap samples to measure impact of a geothermal energy project on the biodiversity of a Costa Rican old-growth rain forest. *Genome*, *63*(9), 407–436.
- Jeunen, G.-J., Knapp, M., Spencer, H. G., Lamare, M. D., Taylor, H. R., Stat, M., Bunce, M., & Gemmell, N. J. (2019). Environmental DNA (eDNA) metabarcoding reveals strong discrimination among diverse marine habitats connected by water movement. *Molecular Ecology Resources*, *19*(2), 426–438.
- Ji, Y., Ashton, L., Pedley, S. M., Edwards, D. P., Tang, Y., Nakamura, A., Kitching, R., Dolman, P. M., Woodcock, P., Edwards, F. A., Larsen, T. H., Hsu, W. W., Benedick, S., Hamer, K. C., Wilcove, D. S., Bruce, C., Wang, X., Levi, T., Lott, M., … Yu, D. W. (2013). Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, *16*(10), 1245–1257.
- Jinbo, U., Kato, T., & Ito, M. (2011). Current progress in DNA barcoding and future implications for entomology. *Entomological Science*, *14*(2), 107–124.
- Johnson, K. P., Adams, R. J., & Clayton, D. H. (2001). Molecular Systematics of Goniodidae (Insecta: Phthiraptera). *Journal of Parasitology*, *87*(4), 862–869.
- Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Molecular Ecology*, *25*(1), 185–202.
- Juen, A., & Traugott, M. (2006). Amplification facilitators and multiplex PCR: Tools to overcome PCR-inhibition in DNA-gut-content analysis of soil-living invertebrates. *Soil Biology and Biochemistry*, *38*(7), 1872–1879.
- Kaartinen, R., Stone, G. N., Hearn, J., Lohse, K., & Roslin, T. (2010). Revealing secret liaisons: DNA barcoding changes our understanding of food webs. *Ecological Entomology*, *35*(5), 623–638.

- Kasparyan, D. R., & López-Ortega, M. (2021). A new species of Ischnus from Mexico (Hymenoptera: Ichneumonidae: Cryptinae: Cryptini). *Zoosystematica Rossica*, *30*(1), 94–100.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649.
- Kemp, J. E., Linder, H. P., & Ellis, A. G. (2017). Beta diversity of herbivorous insects is coupled to high species and phylogenetic turnover of plant communities across short spatial scales in the Cape Floristic Region. *Journal of Biogeography*, *44*(8), 1813–1823.
- Kerley, G. I. H., Landman, M., Ficetola, G. F., Boyer, F., Bonin, A., Rioux, D., Taberlet, P., & Coissac, E. (2018). Diet shifts by adult flightless dung beetles Circellium bacchus, revealed using DNA metabarcoding, reflect complex life histories. *Oecologia*, *188*(1), 107–115.
- Kim, H., & Lee, S. (2008). Molecular systematics of the genus Megoura (Hemiptera: Aphididae) using mitochondrial and nuclear DNA sequences. *Molecules and Cells*, *25*(4), 510–522.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–120.
- Knief, C. (2014). Analysis of plant microbe interactions in the era of next generation sequencing technologies. *Frontiers in Plant Science*, *5*(00). 1-23
- Knox, M. A., Hogg, I. D., Pilditch, C. A., Garcia-R, J. C., Hebert, P. D. N., & Steinke, D. (2020). Contrasting patterns of genetic differentiation for deep-sea amphipod taxa along New Zealand's continental margins. *Deep Sea Research Part I: Oceanographic Research Papers*, *162*, 103323.
- Königsmann, E. (2008). Catalog of Hymenoptera in America North of Mexiko. - Vol. 1, Symphyta and Apocrita (Parasitica) S. 1-1198, 1979. Smithsonian Institution Press Washington, D.C., US $ 30,00. Vol. 2, Apocrita (Acule. *Deutsche Entomologische Zeitschrift*, *28*(1–3), 183–183.
- Korbie, D. J., & Mattick, J. S. (2008). Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature Protocols*, *3*(9), 1452–1456. https://doi.org/10.1038/nprot.2008.133

- Kramer, M. F., & Coen, D. M. (2001). Enzymatic Amplification of DNA by PCR: Standard Procedures and Optimization. In F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, & K. Struhl (Eds.), *Current Protocols in Molecular Biology, 56*, 15.1.1-15.1.14.
- Krehenwinkel, H., Fong, M., Kennedy, S., Huang, E. G., Noriyuki, S., Cayetano, L., & Gillespie, R. (2018). The effect of DNA degradation bias in passive sampling devices on metabarcoding studies of arthropod communities and their associated microbiota. *PLOS ONE*, *13*(1), 1-14.
- Krehenwinkel, H., Graze, M., Rödder, D., Tanaka, K., Baba, Y. G., Muster, C., & Uhl, G. (2016). A phylogeographical survey of a highly dispersive spider reveals eastern Asia as a major glacial refugium for Palaearctic fauna. *Journal of Biogeography*, *43*(8), 1583–1594.
- Krehenwinkel, H., Pomerantz, A., Henderson, J. B., Kennedy, S. R., Lim, J. Y., Swamy, V., Shoobridge, J. D., Graham, N., Patel, N. H., Gillespie, R. G., & Prost, S. (2019). Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *GigaScience*, *8*(5), 1-50.
- Krell, F., T. (2004) Parataxonomy vs. taxonomy in biodiversity studies – pitfalls and applicability of 'morphospecies' sorting. *Biodiversity and Conservation, 13*(5), 795–812
- Kress, W. J., & Erickson, D. L. (2008). DNA Barcoding-a Windfall for Tropical Biology?: Commentary. *Biotropica*, *40*(4), 405–408.
- Krosch, M. N., & Cranston, P. S. (2012). Non-destructive DNA extraction from Chironomidae, including of fragile pupal exuviae, extends analysable collections and enhances vouchering. *CHIRONOMUS Journal of Chironomidae Research*, *25*(00).22-27.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549.
- Lagisz, M., Port, G., & Wolff, K. (2010). A cost-effective, simple and high-throughput method for DNA extraction from insects. *Insect Science 17*, 465–470.
- Lanner, J., Curto, M., Pachinger, B., Neumüller, U., & Meimberg, H. (2019). Illumina midi-barcodes: quality proof and applications. *Mitochondrial DNA Part A*, *30*(3), 490–499.
- Lee, W., Lee, Y., Kim, H., Akimoto, S.-I., & Lee, S. (2014). Developing a new molecular marker for aphid species identification: Evaluation of eleven candidate genes with species-level sampling. *Journal of Asia-Pacific Entomology*, *17*(3), 617–627.

- Lefort, V., Longueville, J., E., & Gascuel, O. (2017). SMS: Smart Model Selection in PhyML. *Molecular Biology and Evolution*, *34*(9), 2422–2424.
- Leggett, R. M., & Clark, M. D. (2017). A world of opportunities with nanopore sequencing. *Journal of Experimental Botany*, *68*(20), 5419–5429.
- Leite, L. A. R. (2012). Mitochondrial pseudogenes in insect DNA barcoding: differing points of view on the same issue. *Biota Neotropica*, *12*(3), 301–308.
- Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, *112*(7), 2076–2081.
- Leray, M., Haenel, Q., & Bourlat, S. J. (2016). Preparation of Amplicon Libraries for Metabarcoding of Marine Eukaryotes Using Illumina MiSeq: The Adapter Ligation Method. In S. J. Bourlat (Ed.), *Marine Genomics* (Vol. 1452, pp. 209–218). Springer New York.
- Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, *10*(1), 1-13.
- Levene, H. (1960). Robust testes for equality of variances. P. 278–292. ***In:*** Contributions to Probability and Statistics. (ed) Olkin I. Stanford University Press, Palo Alto, California.
- Levene, M. J. (2003). Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations. *Science*, *299*(5607), 682–686.
- Levy, S. E., & Boone, B. E. (2019). Next-Generation Sequencing Strategies. *Cold Spring Harbor Perspectives in Medicine*, *9*(7), 1-12.
- Lim, N. K. M., Tay, Y. C., Srivathsan, A., Tan, J. W. T., Kwik, J. T. B., Baloğlu, B., Meier, R., & Yeo, D. C. J. (2016). Next-generation freshwater bioassessment: eDNA metabarcoding with a conserved metazoan primer reveals species-rich and reservoir-specific communities. *Royal Society Open Science*, *3*(11), 166-175.
- Lin, Y.P., Kondo, T., Gullan, P., & Cook, L. G. (2013). Delimiting genera of scale insects: molecular and morphological evidence for synonymising *Taiwansaissetia* Tao, Wong and Chang with *Coccus* Linnaeus (Hemiptera: Coccoidea: Coccidae). *Systematic Entomology*, *38*(2), 249–264.
- Linard, B., Crampton-Platt, A., Moriniere, J., Timmermans, M. J. T. N., Andújar, C., Arribas, P., Miller, K. E., Lipecki, J., Favreau, E., Hunter, A., Gómez-Rodríguez, C.,

Barton, C., Nie, R., Gillett, C. P. D. T., Breeschoten, T., Bocak, L., & Vogler, A. P. (2018). The contribution of mitochondrial metagenomics to large-scale data mining and phylogenetic analysis of Coleoptera. *Molecular Phylogenetics and Evolution*, *128*(00), 1–11.

- Little, D. P. (2014). A DNA mini-barcode for land plants. *Molecular Ecology Resources*, *14*(3), 437–446.
- Liu, M., Clarke, L. J., Baker, S. C., Jordan, G. J., & Burridge, C. P. (2020). A practical guide to DNA metabarcoding for entomological ecologists. *Ecological Entomology*, *45*(3), 373–385.
- Liu, S., Li, Y., Lu, J., Su, X., Tang, M., Zhang, R., Zhou, L., Zhou, C., Yang, Q., Ji, Y., Yu, D. W., & Zhou, X. (2013). SOAPBarcode: revealing arthropod biodiversity through assembly of Illumina shotgun sequences of PCR amplicons. *Methods in Ecology and Evolution*, *4*(12), 1142–1150.
- Loit, K., Adamson, K., Bahram, M., Puusepp, R., Anslan, S., Kiiker, R., Drenkhan, R., & Tedersoo, L. (2019). Relative Performance of MinION (Oxford Nanopore Technologies) versus Sequel (Pacific Biosciences) Third-Generation Sequencing Instruments in Identification of Agricultural and Forest Fungal Pathogens. *Applied and Environmental Microbiology*, *85*(21), 1-20.
- Lorenz, T. C. (2012). Polymerase Chain Reaction: Basic Protocol Plus Troubleshooting and Optimization Strategies. *Journal of Visualized Experiments*, *63*, 3998. https://doi.org/10.3791/3998
- Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D., & Dangl, J. L. (2013). Practical innovations for high-throughput amplicon sequencing. *Nature Methods*, *10*(10), 999–1002.
- Luo, A., Ling, C., Ho, S. Y. W., & Zhu, C.-D. (2018). Comparison of Methods for Molecular Species Delimitation Across a Range of Speciation Scenarios. *Systematic Biology*, *67*(5), 830–846.
- Luz, F. A., Goetz, A. P. M., & Mendonça Jr., M. de S. (2020). Phenotypic matching in ovipositor size in the parasitoid Galeopsomyia sp. (Hymenoptera, Eulophidae) attacking different gall inducers. *Iheringia. Série Zoologia*, *110* (2), e2020008.
- Macher, J. N., Salis, R. K., Blakemore, K. S., Tollrian, R., Matthaei, C. D., & Leese, F. (2016). Multiple-stressor effects on stream invertebrates: DNA barcoding reveals contrasting responses of cryptic mayfly species. *Ecological Indicators*, *61*, 159–169.

- Macher, J.-N., Zizka, V. M. A., Weigand, A. M., & Leese, F. (2018). A simple centrifugation protocol for metagenomic studies increases mitochondrial DNA yield by two orders of magnitude. *Methods in Ecology and Evolution*, *9*(4), 1070–1074.
- Magurran, A. E., Dornelas, M., Moyes, F., & Henderson, P. A. (2019). Temporal β diversity—A macroecological perspective. *Global Ecology and Biogeography*, *28*(12), 1949–1960.
- Mallet, J., & Willmott, K. (2003). Taxonomy: renaissance or Tower of Babel? *Trends in Ecology & Evolution*, *18*(2), 57–59.
- Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods*, *7*(2), 111–118.
- Mardis, E. R. (2017). DNA sequencing technologies: 2006–2016. *Nature Protocols*, *12*(2), 213–218.
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., … Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057), 376–380.
- Marien, A., Debode, F., Aerts, C., Ancion, C., Francis, F., & Berben, G. (2018). Detection of *Hermetia illucens* by real-time PCR. *Journal of Insects as Food and Feed*, *4*(2), 115-122.
- Marsela, M., Hayashida, K., Nakao, R., Chatanga, E., Gaithuma, A. K., Naoko, K., Musaya, J., Sugimoto, C., & Yamagishi, J. (2020). Molecular identification of trypanosomes in cattle in Malawi using PCR methods and nanopore sequencing: epidemiological implications for the control of human and animal trypanosomiases. *Parasite*, *27*(46), 1-10.
- Martoni, F., Valenzuela, I., & Blacket, M. J. (2019). Non-destructive DNA extractions from fly larvae (Diptera: Muscidae) enable molecular identification of species and enhance morphological features. *Austral Entomology*, *58*(4), 848–856.
- Masner, L. (1979). HYMENOPTERA. *Memoirs of the Entomological Society of Canada*, *111*(S108), 485–508.
- McKenna, P., Hoffmann, C., Minkah, N., Aye, P. P., Lackner, A., Liu, Z., Lozupone, C. A., Hamady, M., Knight, R., & Bushman, F. D. (2008). The Macaque Gut Microbiome in Health, Lentiviral Infection, and Chronic Enterocolitis. *PLOS Pathogens*, *4*(2), 1-13.

- McKiernan, H. E., & Danielson, P. B. (2017). Molecular Diagnostic Applications in Forensic Science. In *Molecular Diagnostics* (pp. 371–394). Elsevier.
- Meier, R., Blaimer, B. B., Buenaventura, E., Hartop, E., Rintelen, T., Srivathsan, A., & Yeo, D. (2022). A re-analysis of the data in Sharkey et al.'s (2021) minimalist revision reveals that BINs do not deserve names, but BOLD Systems needs a stronger commitment to open science. *Cladistics*, *38*(2), 264–275.
- Meier, R., Wong, W., Srivathsan, A., & Foo, M. (2016). $1 DNA barcodes for reconstructing complex phenomes and finding rare species in specimen-rich samples. *Cladistics*, *32*(1), 100–110.
- Meier, R., Zhang, G., & Ali, F. (2008). The Use of Mean Instead of Smallest Interspecific Distances Exaggerates the Size of the "Barcoding Gap" and Leads to Misidentification. *Systematic Biology*, *57*(5), 809–813.
- Meiklejohn, K. A., Damaso, N., & Robertson, J. M. (2019). Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. *PLOS ONE*, *14*(6), e0217084.
- Meusnier, I., Singer, G. A., Landry, J.-F., Hickey, D. A., Hebert, P. D., & Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, *9*(1), 1-4.
- Mikheyev, A. S., & Tin, M. M. Y. (2014). A first look at the Oxford Nanopore MinION sequencer. *Molecular Ecology Resources*, *14*(6), 1097–1102.
- Miller, K. E., Hopkins, K., Inward, D. J. G., & Vogler, A. P. (2016). Metabarcoding of fungal communities associated with bark beetles. *Ecology and Evolution*, *6*(6), 1590–1600.
- Miller, S. E., Hausmann, A., Hallwachs, W., & Janzen, D. H. (2016). Advancing taxonomy and bioinventories with DNA barcodes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1702), 1-8.
- Milligan, S. B., Bodeau, J., Yaghoobi, J., Kaloshian, I., Zabel, P., & Williamson, V. M. (1998). The Root Knot Nematode Resistance Gene *Mi* from Tomato Is a Member of the Leucine Zipper, Nucleotide Binding, Leucine-Rich Repeat Family of Plant Genes. *The Plant Cell*, *10*(8), 1307–1319.
- Minas, K., McEwan, N., R., Newbold, C., J., Scott, K., P. (2011) Optimization of a high-throughput CTAB-based protocol for the extraction of qPCR-grade DNA from rumen fluid, plant and bacterial pure cultures. *FEMS Microbiol Lett 325*(2):162–169

- Missa, O., Basset, Y., Alonso, A., Miller, S. E., Curletti, G., De Meyer, M., Eardley, C., Mansell, M. W., & Wagner, T. (2009). Monitoring arthropods in a tropical landscape: relative effects of sampling methods and habitat types on trap catches. *Journal of Insect Conservation*, *13*(1), 103–118.
- Miura, K., Higashiura, Y., & Maeto, K. (2017). Evaluation of easy, non-destructive methods of DNA extraction from minute insects. *Applied Entomology and Zoology, 52*(2), 349–352
- Mockford, E. L. (1997). A New Species of Dicopomorpha (Hymenoptera: Mymaridae) with Diminutive, Apterous Males. *Annals of the Entomological Society of America*, *90*(2), 115–120.
- Monaghan, M. T., Balke, M., Gregory, T. R., & Vogler, A. P. (2005). DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1925–1933.
- Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. B., & Worm, B. (2011). How Many Species Are There on Earth and in the Ocean? *PLoS Biology*, *9*(8), 1-8.
- Mori, A. S., Isbell, F., & Seidl, R. (2018). β-diversity, community assembly, and ecosystem functioning. *Trends in Ecology & Evolution*, *33*(7), 549–564.
- Morinière, J., Araujo, B. C. de, Lam, A. W., Hausmann, A., Balke, M., Schmidt, S., Hendrich, L., Doczkal, D., Fartmann, B., Arvidsson, S., & Haszprunar, G. (2016). Species Identification in Malaise Trap Samples by DNA Barcoding Based on NGS Technologies and a Scoring Matrix. *PLOS ONE*, *11*(5), 1-14.
- Morinière, J., Balke, M., Doczkal, D., Geiger, M. F., Hardulak, L. A., Haszprunar, G., Hausmann, A., Hendrich, L., Regalado, L., Rulik, B., Schmidt, S., Wägele, J., & Hebert, P. D. N. (2019). A DNA barcode library for 5,200 German flies and midges (Insecta: Diptera) and its implications for metabarcoding-based biomonitoring. *Molecular Ecology Resources*, *19*(4), 900–928.
- Mosher, J. J., Bowman, B., Bernberg, E. L., Shevchenko, O., Kan, J., Korlach, J., & Kaplan, L. A. (2014). Improved performance of the PacBio SMRT technology for 16S rDNA sequencing. *Journal of Microbiological Methods*, *104*(00), 59–60.
- Muhammad Tahir, H., & Akhtar, S. (2016). Services of DNA barcoding in different fields. *Mitochondrial DNA Part A*, *27*(6), 4463–4474.
- Murthy, M. K., Khandayataray, P., Tara, M., Buragohain, P., Giri, A., & Gurusubramanian, G. (2022). Optimisation of DNA isolation and PCR techniques for beetle (Order: Coleoptera) specimens. *International Journal of Tropical Insect Science 5*(6). 1-9.

- Musapa, M., Kumwenda, T., Mkulama, M., Chishimba, S., Norris, D. E., Thuma, P. E., & Mharakurwa, S. (2013). A Simple Chelex Protocol for DNA Extraction from Anopheles spp. *Journal of Visualized Experiments*, *71*(3281), 1-6.
- Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M. C., Hirai, A., Takahashi, H., Altaf-Ul-Amin, M., Ogasawara, N., & Kanaya, S. (2011). Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Research*, *39*(13), 1-13.
- New, T. R. (2011). Eric Grissell: Bees, wasps, and ants. The indispensable role of Hymenoptera in gardens: Timber Press, Portland, Oregon and London, 2010, 335 pp, Hardback, £17.99, Au$ 40.00, ISBN 13 9780881929881. *Journal of Insect Conservation*, *15*(4), 609–610.
- Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M., Green, R. E., & Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources*, *18*(5), 927–939.
- Oberprieler, S., Andersen, A., & Moritz, C. (2018). Ants in Australia's Monsoonal Tropics: CO1 Barcoding Reveals Extensive Unrecognised Diversity. *Diversity*, *10*(2), 1-36.
- Oliver, I., & Beattie, A. J. (1993). A Possible Method for the Rapid Assessment of Biodiversity. *Conservation Biology*, *7*(3), 562–568.
- Oppert, B., Stoss, S., Monk, A., & Smith, T. (2019). Optimized Extraction of Insect Genomic DNA for Long-Read Sequencing. *Methods and Protocols*, *2*(4), 1-7.
- Orlofske, J. M., & Baird, D. J. (2013). The tiny mayfly in the room: implications of size-dependent invertebrate taxonomic identification for biomonitoring data properties. *Aquatic Ecology*, *47*(4), 481–494.
- Ortiz, D., & Francke, O. F. (2016). Two DNA barcodes and morphology for multi-method species delimitation in Bonnetina tarantulas (Araneae: Theraphosidae). *Molecular Phylogenetics and Evolution*, *101*(00), 176–193.
- Packer, L., Gibbs, J., Sheffield, C., & Hanner, R. (2009). DNA barcoding and the mediocrity of morphology. *Molecular Ecology Resources*, *9*(1), 42–50.
- Panaccio, M., Georgesz, M., Hollywell, C., & Lew, A. (1993). Direct PCR from solid tissues without DNA extraction. *Nucleic Acids Research*, *21*(19), 46-56.
- Parslow, B. A., Schwarz, M. P., & Stevens, M. I. (2021). Molecular diversity and species delimitation in the family Gasteruptiidae (Hymenoptera: Evanioidea). *Genome*, *64*(3), 253–264.

- Parslow, B. A., Schwarz, M. P., & Stevens, M. I. (2021). Molecular diversity and species delimitation in the family Gasteruptiidae (Hymenoptera: Evanioidea). *Genome*, *64*(3), 253–264.
- Patrick, H. J. H., Chomič, A., & Armstrong, K. F. (2016). Cooled Propylene Glycol as a Pragmatic Choice for Preservation of DNA From Remote Field-Collected Diptera for Next-Generation Sequence Analysis. *Journal of Economic Entomology*, *109*(3), 1469–1473.
- Patzold, F., Zilli, A., & Hundsdoerfer, A. K. (2020). Advantages of an easy-to-use DNA extraction method for minimal-destructive analysis of collection specimens. *PLOS ONE*, *15*(7), 1-24.
- Paydar, S., Ucar, E., Yari, P., Safavi, S. M., Kahrizi, D., Nateqi, M., Mirmoayedi, A., & Yari, K. (2018). A simplified and optimized protocol for total DNA extraction from insect species: applicable for studying genetic diversity and PCR-based specimen identification via partial amplification of cytochrome oxidase I (COI) gene. *Cellular and Molecular Biology*, *64*(12), 22–25.
- Pearson, D. L., Hamilton, A. L., & Erwin, T. L. (2011). Recovery Plan for the Endangered Taxonomy Profession. *BioScience*, *61*(1), 58–63.
- Pease, J., & Sooknanan, R. (2012). A rapid, directional RNA-seq library preparation workflow for Illumina® sequencing. *Nature Methods*, *9*(3), 1-12.
- Pentinsaari, M., Hebert, P. D. N., & Mutanen, M. (2014). Barcoding Beetles: A Regional Survey of 1872 Species Reveals High Identification Success and Unusually Deep Interspecific Divergences. *PLoS ONE*, *9*(9), e108651.
- Peters, R. S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., Kozlov, A., Podsiadlowski, L., Petersen, M., Lanfear, R., Diez, P. A., Heraty, J., Kjer, K. M., Klopfstein, S., Meier, R., Polidori, C., Schmitt, T., Liu, S., Zhou, X., … Niehuis, O. (2017). Evolutionary history of the Hymenoptera. *Current Biology*, *27*(7), 1013–1018.
- Piñol, J., Senar, M. A., & Symondson, W. O. C. (2019). The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology*, *28*(2), 407–419.
- Piper, A. M., Batovska, J., Cogan, N. O. I., Weiss, J., Cunningham, J. P., Rodoni, B. C., & Blacket, M. J. (2019). Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. *GigaScience*, *8*(8), 1-22.

- Pires, A. C., & Marinoni, L. (2010). DNA barcoding and traditional taxonomy unified through Integrative Taxonomy: a view that challenges the debate questioning both methodologies. *Biota Neotropica*, *10*(2), 339–346.
- Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., & Vogler, A. P. (2006). Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. *Systematic Biology*, *55*(4), 595–609.
- Porazinska, D. L., Giblin-Davis, R. M., Faller, L., Farmerie, W., Kanzaki, N., Morris, K., Powers, T. O., Tucker, A. E., Sung, W., & Thomas, W. K. (2009). Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources*, *9*(6), 1439–1450.
- Porco, D., Rougerie, R., Deharvengt, L., & Hebert, P. (2010). Coupling non-destructive DNA extraction and voucher retrieval for small soft-bodied Arthropods in a high throughput context: the example of Collembola. *Molecular Ecology Resources 10*(3), 942–945.
- Porter, T. M., & Hajibabaei, M. (2018). Over 2.5 million COI sequences in GenBank and growing. *PLOS ONE*, *13*(9), 1-16.
- Porter, T. M., & Hajibabaei, M. (2018). Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. *Molecular Ecology*, *27*(2), 313–338.
- Potapov, V., & Ong, J. L. (2017). Examining Sources of Error in PCR by Single-Molecule Sequencing. *PLOS ONE*, *12*(1), e0169774.
- Puillandre, N., Brouillet, S., & Achaz, G. (2021). ASAP: assemble species by automatic partitioning. *Molecular Ecology Resources*, *21*(2), 609–620.
- Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation: ABGD, AUTOMATIC BARCODE GAP DISCOVERY. *Molecular Ecology*, *21*(8), 1864–1877.
- Quick, J., Quinlan, A. R., & Loman, N. J. (2014). A reference bacterial genome dataset generated on the MinION^{TM} portable single-molecule nanopore sequencer. *GigaScience*, *3*(1). 1-6.
- Ranasinghe, U. G. S. L., Eberle, J., Thormann, J., Bohacz, C., Benjamin, S. P., & Ahrens, D. (2022). Multiple species delimitation approaches with *COI* barcodes poorly fit each other and morphospecies – An integrative taxonomy case of Sri Lankan Sericini chafers (Coleoptera: Scarabaeidae). *Ecology and Evolution*, *12*(5).
- Rasnitsyn, A. P. (1988). An Outline of Evolution of the Hymenopterous Insects (Order Vespida). *Oriental Insects*, *22*(1), 115–145.

- Rasool, A., Ahmad, T., Ganai, B. A., & Gull, S. (2018). An overview of molecular identification of insect fauna with special emphasis on chalcid wasps (Hymenoptera: Chalcidoidea) of India. *Acta Agriculturae Slovenica*, *111*(1), 229-243.
- Ratnasingham, S., & Hebert, P. D. N. (2007). BARCODING: bold: The Barcode of Life Data System (http://www.barcodinglife.org): BARCODING. *Molecular Ecology Notes*, *7*(3), 355–364.
- Ratnasingham, S., & Hebert, P. D. N. (2013). A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. *PLoS ONE*, *8*(7), 1-16.
- Renaud, A. K., Savage, J., & Adamowicz, S. J. (2012). DNA barcoding of Northern Nearctic Muscidae (Diptera) reveals high correspondence between morphological and molecular species limits. *BMC Ecology*, *12*(1), 24-36.
- Reuter, J. A., Spacek, D. V., & Snyder, M. P. (2015). High-Throughput Sequencing Technologies. *Molecular Cell*, *58*(4), 586–597.
- Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics*, *13*(5), 278–289.
- Riedel, A., Sagata, K., Suhardjono, Y. R., Tänzler, R., & Balke, M. (2013). Integrative taxonomy on the fast track - towards more sustainability in biodiversity research. *Frontiers in Zoology*, *10*(1), 15-25.
- Ritter, C. D., Häggqvist, S., Karlsson, D., Sääksjärvi, I. E., Muasya, A. M., Nilsson, R. H., & Antonelli, A. (2019). Biodiversity assessments in the 21st century: the potential of insect traps to complement environmental samples for estimating eukaryotic and prokaryotic diversity using high-throughput DNA metabarcoding. *Genome*, *62*(3), 147–159.
- Rivera, J., & Currie, D. C. (2009). Identification of Nearctic black flies using DNA barcodes (Diptera: Simuliidae). *Molecular Ecology Resources*, *9*(1), 224–236.
- Rochlin, I., Santoriello, M. P., Mayer, R. T., Campbell, S. R., (2007). Improved high-throughput method for molecular identification of Culex mosquitoes. *Journal of the American Mosquito Control Association 23* (00), 488-491.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. https://doi.org/10.7717/peerj.2584
- Rohland, N., Glocke, I., Aximu-Petri, A., & Meyer, M. (2018). Extraction of highly degraded DNA from ancient bones, teeth and sediments for high-throughput sequencing. *Nature Protocols*, *13*(11), 2447–2461.

- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, *475*(7356), 348–352.
- Rowley, D. L., Coddington, J. A., Gates, M. W., Norrbom, A. L., Ochoa, R. A., Vandenberg, N. J., & Greenstone, M. H. (2007). Vouchering DNA-barcoded specimens: test of a nondestructive extraction protocol for terrestrial arthropods. *Molecular Ecology Notes*, *7*(6), 915–924.
- Rubbmark, O. R., Sint, D., Horngacher, N., & Traugott, M. (2018). A broadly applicable COI primer pair and an efficient single-tube amplicon library preparation protocol for metabarcoding. *Ecology and Evolution*, *8*(24), 12335–12350.
- Saitoh, S., Aoyama, H., Fujii, S., Sunagawa, H., Nagahama, H., Akutsu, M., Shinzato, N., Kaneko, N., & Nakamori, T. (2016). A quantitative protocol for DNA metabarcoding of springtails (Collembola). Genome. 59(9), 705–723.
- Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425.
- Saleem, S. & Anis, S. B. 2021. New records of Anagrus Haliday (Hymenoptera: Chalcidoidea: Mymaridae) with the description of a new species from India. Munis *Entomology & Zoology*, *16* (1): 124-131.
- Sang, T., & Zhong, Y. (2000). Testing Hybridization Hypotheses Based on Incongruent Gene Trees. *Systematic Biology*, *49*(3), 422–434.
- Sankararaman, H., Manickavasagam, S., & Triapitsyn, S. V. (2022). Four new species of Anagrus Haliday (Hymenoptera: Mymaridae) from India, with a key to Indian species. *Journal of Asia-Pacific Entomology*, *25*(1), 101849-101856.
- Santos, D., Ribeiro, G. C., Cabral, A. D., & Sperança, M. A. (2018). A non-destructive enzymatic method to extract DNA from arthropod specimens: Implications for morphological and molecular studies. *PLOS ONE*, *13*(2), 1-11.
- Recolonisation forestière dans les Basses Laurentides au sud du domaine climacique de l'érablière à bouleau jaune. MSc thesis, University of Montreal, Montreal, Canada
- Savage, C. (2001). *Carte de végétation, Station de biologie des Laurentides, Université de Montréal, 1999-2000*.

- Savolainen, V., Cowan, R. S., Vogler, A. P., Roderick, G. K., & Lane, R. (2005). Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1805–1811.
- Schindel, D. E., & Miller, S. E. (2005). DNA barcoding a useful tool for taxonomists. *Nature*, *435*(7038), 17–17.
- Schirmer, M., D'Amore, R., Ijaz, U. Z., Hall, N., & Quince, C. (2016). Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics*, *17*(1), 1-15.
- Schlick-Steiner, B. C., Steiner, F. M., Seifert, B., Stauffer, C., Christian, E., & Crozier, R. H. (2010). Integrative Taxonomy: A Multisource Approach to Exploring Biodiversity. *Annual Review of Entomology*, *55*(1), 421–438.
- Schmidt, K., Mwaigwisya, S., Crossman, L. C., Doumith, M., Munroe, D., Pires, C., Khan, A. M., Woodford, N., Saunders, N. J., Wain, J., O'Grady, J., & Livermore, D. M. (2017). Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *Journal of Antimicrobial Chemotherapy*, *72*(1), 104–114.
- Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, *15*(6), 1289–1303.
- See Kemler, M., Garnas, J., Wingfield, M. J., Gryzenhout, M., Pillay, K.-A., & Slippers, B. (2013). Ion Torrent PGM as Tool for Fungal Community Analysis: A Case Study of Endophytes in Eucalyptus grandis Reveals High Taxonomic Diversity. *PLoS ONE*, *8*(12), 1-9.
- Seibold, S., Gossner, M. M., Simons, N. K., Blüthgen, N., Müller, J., Ambarlı, D., Ammer, C., Bauhus, J., Fischer, M., Habel, J. C., Linsenmair, K. E., Nauss, T., Penone, C., Prati, D., Schall, P., Schulze, E.-D., Vogt, J., Wöllauer, S., & Weisser, W. W. (2019). Arthropod decline in grasslands and forests is associated with landscape-level drivers. *Nature*, *574*(7780), 671–674.
- Sellers, G. S., Muri, C. D., Gómez, A., & Hänfling, B. (2018). Mu-DNA: a modular universal DNA extraction method adaptable for a wide range of sample types. *Metabarcoding and Metagenomics*, *2*(56), 1-7.
- Sharanowski, B. J., Ridenbaugh, R. D., Piekarski, P. K., Broad, G. R., Burke, G. R., Deans, A. R., Lemmon, A. R., Moriarty Lemmon, E. C., Diehl, G. J., Whitfield, J. B., & Hines, H.

M. (2021). Phylogenomics of Ichneumonoidea (Hymenoptera) and implications for evolution of mode of parasitism and viral endogenization. *Molecular Phylogenetics and Evolution*, *156*(5), 107-123.

- Sharkey, M. J., Carpenter, J. M., Vilhelmsen, L., Heraty, J., Liljeblad, J., Dowling, A. P. G., Schulmeister, S., Murray, D., Deans, A. R., Ronquist, F., Krogmann, L., & Wheeler, W. C. (2012). Phylogenetic relationships among superfamilies of Hymenoptera. *Cladistics*, *28*(1), 80–112.
- Sheffield, C. S., Hebert, P. D. N., Kevan, P. G., & Packer, L. (2009). DNA barcoding a regional bee (Hymenoptera: Apoidea) fauna and its potential for ecological studies. *Molecular Ecology Resources*, *9*, 196–207.
- Sheikh, A.H., Thomas, M., Bhandari, R., Bunkar, K. (2016). Light trap and insect sampling: an overview. *Int J Curr Res, 8(11)*, 40868–40873.
- Shekhovtsov, S. V., Shekhovtsova, I. N., & Peltek, S. E. (2019). DNA Barcoding: Methods and Approaches. *Biology Bulletin Reviews*, *9*(6), 475–483.
- Shokralla, S., Gibson, J. F., Nikbakht, H., Janzen, D. H., Hallwachs, W., & Hajibabaei, M. (2014). Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Molecular Ecology Resources*, *14*(5), 892–901.
- Simon, S., Schierwater, B., & Hadrys, H. (2010). On the value of Elongation factor-1α for reconstructing pterygote insect phylogeny. *Molecular Phylogenetics and Evolution*, *54*(2), 651–656.
- Singer, E., Bushnell, B., Coleman-Derr, D., Bowman, B., Bowers, R. M., Levy, A., Gies, E. A., Cheng, J.-F., Copeland, A., Klenk, H.-P., Hallam, S. J., Hugenholtz, P., Tringe, S. G., & Woyke, T. (2016). High-resolution phylogenetic microbial community profiling. *The ISME Journal*, *10*(8), 2020–2032.
- Singer, G. A. C., Fahner, N. A., Barnes, J. G., McCarthy, A., & Hajibabaei, M. (2019). Comprehensive biodiversity analysis via ultra-deep patterned flow cell technology: a case study of eDNA metabarcoding seawater. *Scientific Reports*, *9*(1), 1-12.
- Sinha, R., Stanley, G., Gulati, G. S., Ezran, C., Travaglini, K. J., Wei, E., Chan, C. K. F., Nabhan, A. N., Su, T., Morganti, R. M., Conley, S. D., Chaib, H., Red-Horse, K., Longaker, M. T., Snyder, M. P., Krasnow, M. A., & Weissman, I. L. (2017). Index switching causes "spreading-of-signal" among multiplexed samples in Illumina HiSeq 4000 DNA sequencing. *BioRxiv*, *9*(00), 1-29.

- Siqueira, J. F., Fouad, A. F., & Rôças, I. N. (2012). Pyrosequencing as a tool for better understanding of human microbiomes. *Journal of Oral Microbiology*, *4*(1), 1-15.
- Smith, M. A., Bertrand, C., Crosby, K., Eveleigh, E. S., Fernandez-Triana, J., Fisher, B. L., Gibbs, J., Hajibabaei, M., Hallwachs, W., Hind, K., Hrcek, J., Huang, D.-W., Janda, M., Janzen, D. H., Li, Y., Miller, S. E., Packer, L., Quicke, D., Ratnasingham, S., … Zhou, X. (2012). Wolbachia and DNA Barcoding Insects: Patterns, Potential, and Problems. *PLoS ONE*, *7*(5), e36514. https://doi.org/10.1371/journal.pone.0036514
- Smith, M. A., Fisher, B. L., & Hebert, P. D. N. (2005). DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1825–1834.
- Sonet, G., Jordaens, K., Braet, Y., Bourguignon, L., Dupont, E., Backeljau, T., de Meyer, M., & Desmyter, S. (2013). Utility of GenBank and the Barcode of Life Data Systems (BOLD) for the identification of forensically important Diptera from Belgium and France. *ZooKeys*, *365*, 307–328.
- Spiess, A.-N., Mueller, N., & Ivell, R. (2004). Trehalose Is a Potent PCR Enhancer: Lowering of DNA Melting Temperature and Thermal Stabilization of Taq Polymerase by the Disaccharide Trehalose. *Clinical Chemistry*, *50*(7), 1256–1259.
- Srivathsan, A., Ang, A., Vogler, A. P., & Meier, R. (2016). Fecal metagenomics for the simultaneous assessment of diet, parasites, and population genetics of an understudied primate. *Frontiers in Zoology*, *13*(1), 17.
- Srivathsan, A., Hartop, E., Puniamoorthy, J., Lee, W. T., Kutty, S. N., Kurina, O., & Meier, R. (2019). Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. *BMC Biology*, *17*(1), 1-20.
- Srivathsan, A., Lee, L., Katoh, K., Hartop, E., Kutty, S. N., Wong, J., Yeo, D., & Meier, R. (2021). ONTbarcoder and MinION barcodes aid biodiversity discovery and identification by everyone, for everyone. *BMC Biology*, *19*(1), 217.
- Staats, M., Arulandhu, A. J., Gravendeel, B., Holst-Jensen, A., Scholtens, I., Peelen, T., Prins, T. W., & Kok, E. (2016). Advances in DNA metabarcoding for food and wildlife forensic species identification. *Analytical and Bioanalytical Chemistry*, *408*(17), 4615–4630.

- Ståhlberg, A., Krzyzanowski, P. M., Egyud, M., Filges, S., Stein, L., & Godfrey, T. E. (2017). Simple multiplexed PCR-based barcoding of DNA for ultrasensitive mutation detection by next-generation sequencing. *Nature Protocols*, *12*(4), 664–682.
- Stahlhut, J. K., Fernández-Triana, J., Adamowicz, S. J., Buck, M., Goulet, H., Hebert, P. D., Huber, J. T., Merilo, M. T., Sheffield, C. S., Woodcock, T., & Smith, M. A. (2013). DNA barcoding reveals diversity of Hymenoptera and the dominance of parasitoids in a sub-arctic environment. *BMC Ecology*, *13*(1), 2. 1-13.
- Stork, N. E. (2018). How Many Species of Insects and Other Terrestrial Arthropods Are There on Earth? *Annual Review of Entomology*, *63*(1), 31–45. https://doi.org/10.1146/annurev-ento-020117-043348
- Stribling, J. B., Pavlik, K. L., Holdsworth, S. M., & Leppo, E. W. (2008). Data quality, performance, and uncertainty in taxonomic identification for biological assessments. *Journal of the North American Benthological Society*, *27*(4), 906–919.
- Strutzenberger, P., Brehm, G., & Fiedler, K. (2011). DNA barcoding-based species delimitation increases species count of Eois (Geometridae) moths in a well-studied tropical mountain forest by up to 50%. *Insect Science*, *18*(3), 349–362.
- Suaste-Dzul, A. P., Rodríguez-Vélez, J. M., Rodríguez-Vélez, B., Arredondo-Bernal, H. C., & Gallou, A. (2019). Non-destructive DNA extraction methods for entomophagous insects with emphasis on biological control. *Genome*, *62*(4), 287–293.
- Sultana, S., Ali, Md. E., Hossain, M. A. M., Asing, Naquiah, N., & Zaidul, I. S. M. (2018). Universal mini COI barcode for the identification of fish species in processed products. *Food Research International*, *10*(2), 19–28.
- Sun, W. (2010). Chapter 4 - Nucleic Extraction and Amplification. In W. W. Grody, R. M. Nakamura, C. M. Strom, & F. L. Kiechle (Eds.), *Molecular Diagnostics* (pp. 35–47). Academic Press.
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050.
- Taekul, C., Valerio, A. A., Austin, A. D., Klompen, H., & Johnson, N. F. (2014). Molecular phylogeny of telenomine egg parasitoids (Hymenoptera: Platygastridae s.l.: Telenominae): evolution of host shifts and implications for classification: Molecular phylogeny of telenomine egg parasitoids. *Systematic Entomology*, *39*(1), 24–35.

- Tahir, H. M., Noor, A., Mehmood, S., Sherawat, S. M., & Qazi, M. A. (2018). Evaluating the accuracy of morphological identification of insect pests of rice crops using DNA barcoding. *Mitochondrial DNA Part B*, *3*(2), 1220–1224.
- Takahara, T., Minamoto, T., & Doi, H. (2013). Using Environmental DNA to Estimate the Distribution of an Invasive Fish Species in Ponds. *PLOS ONE*, *8*(2), 1-5.
- Talavera, G., Dincă, V., & Vila, R. (2013). Factors affecting species delimitations with the GMYC model: insights from a butterfly survey. *Methods in Ecology and Evolution*, *4*(12), 1101–1110.
- Tang, M., Hardman, C. J., Ji, Y., Meng, G., Liu, S., Tan, M., Yang, S., Moss, E. D., Wang, J., Yang, C., Bruce, C., Nevard, T., Potts, S. G., Zhou, X., & Yu, D. W. (2015). High-throughput monitoring of wild bee diversity and abundance via mitogenomics. *Methods in Ecology and Evolution*, *6*(9), 1034–1043.
- Taub, M. A., Corrada Bravo, H., & Irizarry, R. A. (2010). Overcoming bias and systematic errors in next generation sequencing data. *Genome Medicine*, *2*(12), 87.
- Tedersoo, L., Tooming-Klunderud, A., & Anslan, S. (2018). PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytologist*, *217*(3), 1370–1385.
- Théry, T., Brockerhoff, E. G., Carnegie, A. J., Chen, R., Elms, S. R., Hullé, M., Glatz, R., Ortego, J., Qiao, G.-X., Turpeau, É., & Favret, C. (2017). EF-1α DNA Sequences Indicate Multiple Origins of Introduced Populations of Essigella californica (Hemiptera: Aphididae). *Journal of Economic Entomology*, *110*(3), 1269–1274.
- Tortorici, F., Talamas, E. J., Moraglio, S. T., Pansa, M. G., Asadi-Farfar, M., Tavella, L., & Caleca, V. (2019). A morphological, biological and molecular approach reveals four cryptic species of Trissolcus Ashmead (Hymenoptera, Scelionidae), egg parasitoids of Pentatomidae (Hemiptera). *Journal of Hymenoptera Research*, *73*, 153–200.
- Truett, G. E., Heeger, P., Mynatt, R. L., Truett, A. A., Walker, J. A., & Warman, M. L. (2000). Preparation of PCR-Quality Mouse Genomic DNA with Hot Sodium Hydroxide and Tris (HotSHOT). *BioTechniques*, *29*(1), 52–54.
- Tuomisto, H. (2010). A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography*, *33*(1), 2–22.

- Ulmer, J., M., Mikón I., Deans, A., R., Krogmann, L. (2021) The Waterston's evaporatorium of Ceraphronidae (Ceraphronoidea, Hymenoptera): A morphological barcode to a cryptic taxon. *Journal of Hymenoptera Research 85*(4): 29-56.
- Uroz, S., Buée, M., Murat, C., Frey-Klett, P., & Martin, F. (2010). Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environmental Microbiology Reports*, *2*(2), 281–288.
- van Achterberg, K., Taeger, A., Blank, S., Zwakhals, K., Viitasaari, M., Yu, D., & de Jong, Y. (2017). Fauna Europaea: Hymenoptera – Symphyta & Ichneumonoidea. *Biodiversity Data Journal*, *5*, e14650.
- Vasilita, C., Moser, M., & Krogmann, L. (2022). Mission possible: an optimised protocol for the unbarcodable Ceraphronoidea (Hymenoptera). *Biodiversity Data Journal*, *10*, e84860.
- Veijalainen, A., Sääksjärvi, I. E., Erwin, T. L., Gómez, I. C., & Longino, J. T. (2013). Subfamily composition of Ichneumonidae (Hymenoptera) from western Amazonia: Insights into diversity of tropical parasitoid wasps: *Insights into tropical parasitoid diversity*. *Insect Conservation and Diversity*, *6*(1), 28–37.
- Vierna, J., Doña, J., Vizcaíno, A., Serrano, D., & Jovani, R. (2017). PCR cycles above routine numbers do not compromise high-throughput DNA barcoding results. *Genome*, *60*(10), 868–873.
- Vinarski, M. V. (2020). Roots of the taxonomic impediment: Is the "integrativeness" a remedy? *Integrative Zoology*, *15*(1), 2–15.
- Vogler, A. P., & Monaghan, M. T. (2007). Recent advances in DNA taxonomy. *Journal of Zoological Systematics and Evolutionary Research*, *45*(1), 1–10.
- Wagner, D. L. (2020). Insect Declines in the Anthropocene. *Annual Review of Entomology*, *65*(1), 457–480.
- Wagner, J., Coupland, P., Browne, H. P., Lawley, T. D., Francis, S. C., & Parkhill, J. (2016). Evaluation of PacBio sequencing for full-length bacterial 16S rRNA gene classification. *BMC Microbiology*, *16*(1), 1-16.
- Walder, F., Schlaeppi, K., Wittwer, R., Held, A. Y., Vogelgsang, S., & van der Heijden, M. G. A. (2017). Community Profiling of Fusarium in Combination with Other Plant-Associated Fungi in Different Crop Species Using SMRT Sequencing. *Frontiers in Plant Science*, *8*(00), 1-17.

- Walsh, P. S., Metzger, D. A., & Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, *10*(4), 506–513.
- Wang, W. Y., Srivathsan, A., Foo, M., Yamane, S. K., & Meier, R. (2018). Sorting specimen-rich invertebrate samples with cost-effective NGS barcodes: Validating a reverse workflow for specimen processing. *Molecular Ecology Resources*, *18*(3), 490–501.
- Wang, Y., Yang, Q., & Wang, Z. (2015). The evolution of nanopore sequencing. *Frontiers in Genetics*, *5*(00), 1-21
- Wang, Y., Zhou, Q.-S., Qiao, H.-J., Zhang, A.-B., Yu, F., Wang, X.-B., Zhu, C.-D., & Zhang, Y.-Z. (2016). Formal nomenclature and description of cryptic species of the Encyrtus sasakii complex (Hymenoptera: Encyrtidae). *Scientific Reports*, *6*(1), 343-372.
- Wang, Y.-S., Dai, T.-M., Tian, H., Wan, F.-H., & Zhang, G.-F. (2019). Comparative analysis of eight DNA extraction methods for molecular research in mealybugs. *PLOS ONE*, *14*(12), 1-15.
- Weirather, J. L., de Cesare, M., Wang, Y., Piazza, P., Sebastiano, V., Wang, X.-J., Buck, D., & Au, K. F. (2017). Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research*, *6*(100), 1-32.
- Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A. G., Alt, K. W., Caramelli, D., Dresely, V., Farrell, M., Farrer, A. G., Francken, M., Gully, N., Haak, W., Hardy, K., Harvati, K., Held, P., Holmes, E. C., Cooper, A. (2017). Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature*, *544*(7650), 357–361.
- Wheeler, Q. D. (2004). Taxonomy: Impediment or Expedient? *Science*, *303*(5656), 285–285.
- Winand, R., Bogaerts, B., Hoffman, S., Lefevre, L., Delvoye, M., Van Braekel, J., Fu, Q., Roosens, N. H., De Keersmaecker, S. C., & Vanneste, K. (2019). Targeting the 16S rRNA Gene for Bacterial Identification in Complex Mixed Samples: Comparative Evaluation of Second (Illumina) and Third (Oxford Nanopore Technologies) Generation Sequencing Technologies. *International Journal of Molecular Sciences*, *21*(1), 1-22.
- Willis, F. R. S. (1925). A mathematical theory of evolution, based on the conclusions. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, *213*(402–410), 21–87.

- Wong, W. H., Tay, Y. C., Puniamoorthy, J., Balke, M., Cranston, P. S., & Meier, R. (2014). 'Direct PCR' optimization yields a rapid, cost-effective, nondestructive and efficient method for obtaining DNA barcodes without DNA extraction. *Molecular Ecology Resources*, *14*(6), 1271–1280.
- Yang, C., Tan, S., Meng, G., Bourne, D. G., O'Brien, P. A., Xu, J., Liao, S., Chen, A., Chen, X., & Liu, S. (2018). Access COI barcode efficiently using high throughput Single-End 400 bp sequencing. *BioRxiv*, *49*(8618). 1-17.
- Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences*, *107*(20), 9264–9269.
- Yang, Z., & Rannala, B. (2014). Unguided Species Delimitation Using DNA Sequence Data from Multiple Loci. *Molecular Biology and Evolution*, *31*(12), 3125–3135. https://doi.org/10.1093/molbev/msu279
- Yeo, D., Puniamoorthy, J., Ngiam, R. W. J., & Meier, R. (2018). Towards holomorphology in entomology: rapid and cost-effective adult-larva matching using NGS barcodes: Life-history stage matching with NGS barcodes. *Systematic Entomology*, *43*(4), 678–691.
- Yeo, D., Srivathsan, A., & Meier, R. (2019). Mini-barcodes are equally useful for species identification and more suitable for large-scale species discovery in Metazoa than full-length barcodes. *BioRxiv*, *14*(8), 1-48.
- Yeo, D., Srivathsan, A., & Meier, R. (2020). Longer is Not Always Better: Optimizing Barcode Length for Large-Scale Species Discovery and Identification. *Systematic Biology*, *69*(5), 999–1015.
- Yoshimoto, C., M. (1984). The insects and arachnids of Canada. Part 12. The families and subfamilies of Canadian chalcidoid wasps. Hymenoptera: Chalcidoidea (No. 1760).
- Yu, D. S. K., Achterberg van, C., Horstmann, K. (2016) Taxapad 2016, Ichneumonoidea 2015. Database on flash-drive. www.taxapad.com, Nepean, Ontario, Canada.
- Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring: Biodiversity soup. *Methods in Ecology and Evolution*, *3*(4), 613–623.
- Yu, H.J., & You, Z.H. (2010). Comparison of DNA Truncated Barcodes and Full-Barcodes for Species Identification. In D.-S. Huang, X. Zhang, C. A. Reyes García, & L. Zhang (Eds.), *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*, (pp. 108–114). 0-14.

- Zeale, M. R. K., Butlin, R. K., Barker, G. L. A., Lees, D. C., & Jones, G. (2011). Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces: DNA BARCODING. *Molecular Ecology Resources*, *11*(2), 236–244.
- Zepeda-Mendoza, M. L., Bohmann, K., Carmona Baez, A., & Gilbert, M. T. P. (2016). DAMe: a toolkit for the initial processing of datasets with PCR replicates of double-tagged amplicons for DNA metabarcoding analyses. *BMC Research Notes*, *9*(1), 1-13.
- Zhang, Z. (2011). Animal biodiversity: An introduction to higher-level classification and taxnomic richness. *Zootaxa, 3148* (1), 7–12.
- Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, *29*(22), 2869–2876.
- Zhou, X., Li, Y., Liu, S., Yang, Q., Su, X., Zhou, L., Tang, M., Fu, R., Li, J., & Huang, Q. (2013). Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience*, *2*(1), 1-4.
- Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., Chariton, A. A., Creer, S., Coissac, E., Deagle, B. E., Barba, M. D., Dickie, I. A., Dumbrell, A. J., Ficetola, G. F., Fierer, N., Fumagalli, L., Gilbert, M. T. P., Jarman, S., Jumpponen, A., … Taberlet, P. (2019). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, *28*(8), 1857–1862.

Electronic references

- Austin, A., & Dowton, M. (2000). *Hymenoptera: Evolution, Biodiversity and Biological Control: Evolution, Biodiversity and Biological Control*. Csiro Publishing. Available at: https://books.google.ca/books?hl=fr&lr=&id=V09L18iVQ7gC&oi=fnd&pg=PA3&dq=Austin+and+Dowton,+2000*&ots=PyoztmhRZK&sig=r3MyWwQmg0Ia5LiRyLxYLJdYc4g&redir_esc=y#v=onepage&q=Austin%20and%20Dowton%2C%202000*&f=false.
- Berman, J., J. (2016). in Data Simplification
- Chen, H. (2018). Systematics and evolution of the superfamily Platygastroidea (Insecta: Hymenoptera). ProQuest Dissertations & Theses Global. 404. Available at: file:///C:/Users/Malek/Desktop/master/Donn%C3%A9e%20master/Delimitation%20methods%20used/ASaP,%20ABGD,%20bPTP%20important/out.pdf
- Council of Canadian Academies & Expert Panel on Biodiversity Science (2010) Canadian taxonomy: exploring biodiversity, creating opportunity. Ottawa, Canada. Available at: https://www.natureserve.org/biodiversity
- Genome Québec. (2020). Available from: https://www.scienceexchange.com/labs/centre-d-expertise-et-de-service-genome-quebec
- Goulet, H., & Huber, J. T. (Eds.). (1993). *Hymenoptera of the world: an identification guide to families*. Centre for Land and Biological Resources Research. Available at: http://www.esc-sec.ca/aafcmonographs/hymenoptera_of_the_world.pdf
- Hoagland, E. (1995). The Taxonomic Impediment and the Convention on Biodiversity. Available at: http:// flyaqis.museum.vic. gov.au/chaec/taximp.html
- Michonneau, F. (2017). *Using Gmyc For Species Delineation*. Available at: https://doi.org/10.5281/ZENODO.838260
- Porter, T. M., & Hajibabaei, M. (2017). *Automated high throughput animal DNA metabarcode classification* (p. 219675). bioRxiv. Available at: https://www.biorxiv.org/content/10.1101/219675v1
- Qiagen DNeasy Blood and Tissue Handbook. (2020). Available at: file:///C:/Users/Compu-Med/Downloads/HB-2061 003_HB_DNY_Blood_Tissue_0720_WW.pdf
- QuickExtract™ DNA Extraction Solution. (2018). Available at: https://www.lucigen.com/docs/manuals/MA150E-QuickExtract-DNA-Solution.pdf
- R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

- Saunders, T. E. (2018). *Optimising Capture Methods for the Evaluation of Parasitoid Wasp Diversity* [Preprint]. Thesis Commons. Available at : https://doi.org/10.31237/osf.io/gqpyf
- The MasterPure™ Complete DNA and RNA Purification Kit. (2012). Available at: https://www.lucigen.com/docs/manuals/masterpure-complete-dna-and-rna-purification-kit-bulk-reagents.pdf

Addenda

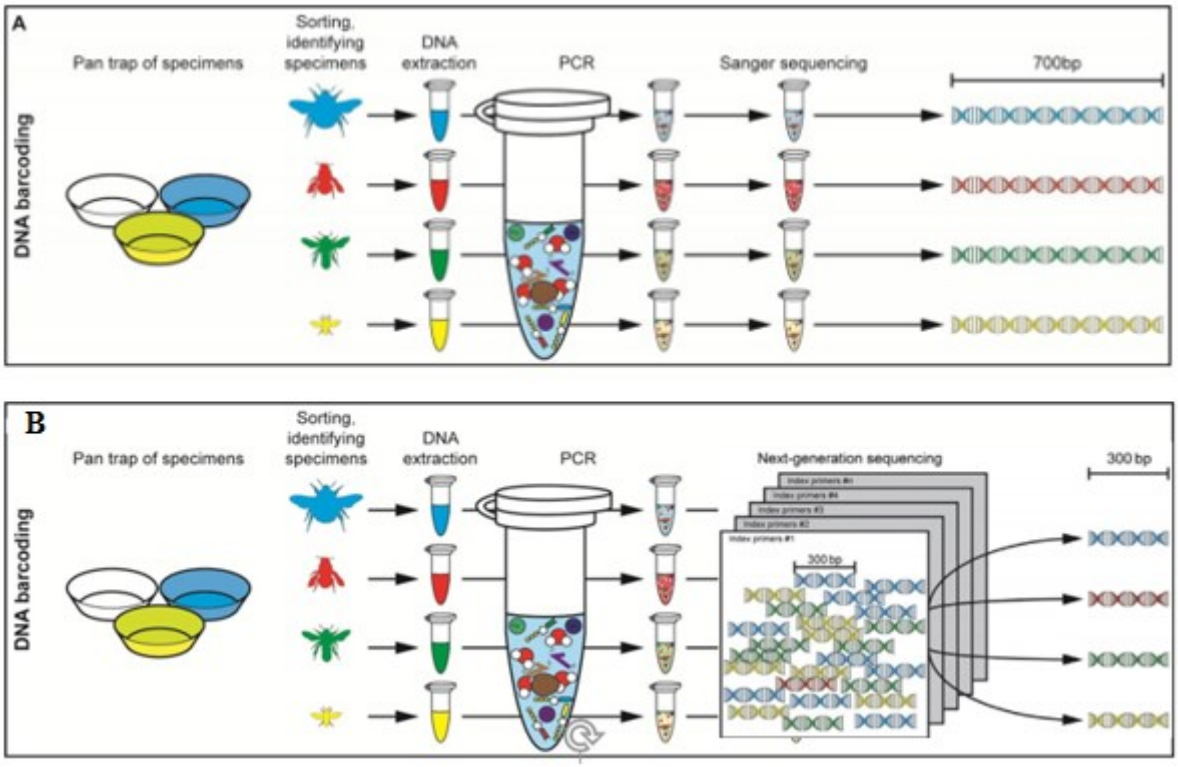


Figure 1. DNA barcoding and meta-barcoding. **A:** specimens and all downstream processes must be kept separate. The DNA from each representative specimen is extracted separately, amplified by PCR at specific loci and Sanger sequenced, to produce a reference database.
B: DNA barcoding protocol combined with HTS sequencing (modified from Gill *et al*., 2016), called Sample multiplexing or DNA meta-barcoding.

Table 1. Non- destructive DNA extraction methods

DNA extraction chemistry	Methods	Advantages	Disadvantages	References
Phenol/chloroform extraction	Modifiable phenol/chloroform extraction method	- Preserve cuticle intact. - Clear specimens easily. - Reliable.	-Toxic. - Not easy. - Time consuming (Overnight incubation (+ 24hr). -Cuticle discoloration	(Favret and Voegtlin, 2004) (Favret, 2005)
	Protocol with CaCl2 buffer	-Efficient (for larger insects or those heavily sclerotized). - Reliable (For old specimens +23 years old). - Cheap.	- Time consuming (+18hr). -Toxic. -DNA contamination risk -Cause exoskeleton discoloration.	(Gilbert *et al*., 2007) (Guzmán-Larralde *et al*., 2017)
	Cetyltrimethyl ammonium bromide method (CTAB)	- High DNA yield. -Cheap (0.63–0.87USD).	-Time-consuming. -Toxic. - Preserve inhibitors.	(Chen *et al*., 2010) (Nancy *et al*., 2010)

	The sodium dodecyl sulfate method (SDS)	-High DNA yield. -Cheap (0.62–0.86USD).	-Time-consuming. -Toxic. - Destructive	(Chen *et al*., 2010) (Wang *et al*., 2019)
Salting out Extraction	The MasterPure™ Complete DNA and RNA Purification Kit	- Fast (30min) - Not toxic -Inexpensive	-May preserve inhibitors	(Evans *et al*., 2013) (The MasterPure™ Complete DNA and RNA Purification Kit, 2012)
Spin-column of DNA-binding membrane extraction	DNeasy® Blood & Tissue (QIAGEN ©).	-Safe and easy. -Reliable -Useful for all insects. - Reduce the time for slide preparations.	- Relatively expensive. -Tedious and labor intensive. -Can be time consuming (2hr-22hr).	(Hernandez *et al*., 2012) (Bahder *et al*., 2015) (Giantsis *et al*., 2015) (Guzmán-Larralde *et al*., 2017) (Miura *et al*., 2017) (Théry *et al*., 2017)

				(Santos *et al*., 2018) (Martoni *et al*., 2019) (Suaste *et al.,* 2019) (Velasco-Cuervo *et al*., 2019)
	PCR & DNA Clean-up Kit (Biolabs, New England).	-Easy. -Short. -Extract DNA from old specimens.	- Minimal destructive method (leg) - Relatively expensive.	(Patzold *et al*., 2020)
	E.Z.N.A. Insect DNA Kit (Omega BioTek, Norcross, GA, USA)	- Higher quality and longer DNA - Easy -Rapid (-3hr)	- Minimal destruction. -Expensive ($2.29 to $3.60 per sample).	(Oppert *et al*., 2019)
Resin Chelation extraction	Chelex extraction	- Low cost - Fast (37 min)	- Sensitive to PCR inhibitors	(Hoff-Olsen *et al*.,1999)

		-Reliable -Low toxicity	-May require minimal destruction (leg)	(Musapa *et al*., 2013) (Miura *et al*., 2017)
Heat treatment (incubation) extraction	QuickExtract™ (Lucigen)	- Fast -Simple -Inexpensive -Not toxic -Easily used for large scale		(Srivathsan *et al*., 2019).
	PrepGEM Technique	- Easy -Fast (20min) -Cheap -Reliable		(Asghar *et al*., 2015)
	HotShot	- Easy -Fast (40 min) -Cheap	-Produce an exoskeleton discoloration in the smaller species.	(Truett *et al*., 2000). (Guzmán-Larralde *et al*., 2017) (Suaste *et al*., 2019)

		- Low toxicity -Efficient in recovery of a full length DNA sequence.		
Magnetic bead technique	ChargeSwitch® Forensic DNA Purification Kit	-Simple. -Fast (25min) -Clean -Not toxic - High- quality DNA	-Sensitive - Slightly lower yield	(Asghar *et al*., 2015)
	A Medics' Magsi DNA Vegetal kit	-High throughput, robotic, liquid handling processes. - Cost effective. -Fast (2hr). -Cheap.	-Minimal destruction (leg)	(Lanner *et al*., 2019)

Table 2. Comparison of available HTS platforms

NGS technologies	Read length	Sequencing cost per sample	Run time (hours)	Advantages	Disadvantages	References
454/Roche pyrosequencing	400bp (single end)	$10	24 hr	-Precision: 99,9% - Simple -Faster throughout -Added information content	-Error of heteroplasmic sequences - Read length restrictions -Confusion with sequences from intracellular endosymbiotic bacteria	(Siqueira*et al*., 2012) (Shokralla *et al*., 2014) (Hsieh *et al*., 2020)
Illumina	150–300bp (paired end)	$5 to $150	1 to 11 days	- Precision: 99.9% -Simple - The overall error rates are below 1% - Analysis of >1000 samples in a single run at sufficient sequencing depth -Greater throughput	- Error of substitution - Read length restrictions - Subject to GC bias -Requires high concentrations of DNA -Generation of self-chimeric sequences	(Bentley *et al*., 2008). (Guo *et al*., 2008). (Dohm*et al*., 2008). (Reuter *et al*., 2015)

						(Meier *et al*., 2016) (Leggett et Clark, 2017) (Tedersoo *et al*., 2018) (Lanner *et al*., 2019) (Piper *et al*., 2019)
Illumina 'MiSeq sequencer'	300bp (paired end)	<$50	56hr	- Fast - Low instrument cost - Produces up to 25 million paired-end -Low sequencing error rates -Well-established bioinformatic procedures	-More expensive	
Illumina 'HiSeq sequencer'	150bp (paired end)	<$10	84hr	- Run in rapid mode - Increases throughput	-Necessitating substantial logistical efforts - High instrument cost	
Illumina 'NextSeq sequencer'	150bp (paired end)	<$15	Less than 30hr	-Reduces data processing times -Reduces cost -Increases throughput	-Necessitating substantial logistical efforts	

Ion Torrent	200–400bp (single end)	$66.8-$950	2–8hr	- Precision: 98% - Cheap equipment -Fast	- Error of Insertions and deletion - Fluctuating sequence quality - Read length restrictions	(Rothberg *et al*., 2011) (Reuter *et al*., 2015) (Tedersoo *et al*., 2018).
Pacific Biosciences Single-molecule real-time (SMRT)	100 kb Up to 40kb (single end or circular consensus)	<$15	15hr	- Long readings -Fast -Inexpensive - Much less sensitive to GC sequence content -PCR amplification can be skiped	-High error right (11%) -Moderate precision: 87% -Expensive Equipment - Moderate throughtput	(Levene *et al*., 2003). (Eid *et al*., 2009). (Reuter *et al*., 2015) (Loit *et al*., 2019) (Tedersoo *et al*., 2018).

Oxford Nanopore Technologies' \	Variable: depends on library preparation (1D or 2D reads): Average read lengths over 10 kb on a single flow cell ~2 Mb	<$25	1–72 hr	-Precision: > 99.5% - Reasonable cost - Lightweight -Fast -Impressive data output -Does not require sophisticated skills in biological research -PCR amplification can be skiped	-Higher raw read error rate (12-22 %) -Very high run failure rate.	(Quick *et al*., 2014). (Wang *et al*., 2015) (Reuter *et al*., 2015) (Marsela *et al*., 2020) (Loit *et al*., 2019) (Srivathsan *et al*., 2019).