Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Department of Ecology

Community phylogenetics in the era of DNA-barcoding: potentials and challenges

Johanna Orsholm

Introductory Research Essay
Department of Ecology
SLU

Uppsala 2024

Contents

Abstract

The evolutionary history of taxa in a community reflects the ecological processes which have been particularly crucial in shaping the communities of today. Integrating phylogenetic relationships among taxa in ecological analysis thus offer valuable insight into the underlying mechanisms governing community assembly and biodiversity patterns. Recent advancement in molecular techniques have drastically popularized a method known as DNA barcoding, where a short genetic marker is sequenced. For animals, the most common marker is the COI barcode region. The primary use of COI barcodes is taxonomic classification, but much potential remains untapped in deriving community phylogenetic information from this data. However, rapid saturation and strong functional constraints limits the phylogenetic information of the COI barcode, especially for older divergences. Inferring phylogenetic relationships thus become challenging for taxonomic groups older than a few million years. To improve phylogenetic inference from barcodes, additional information can be included to guide the analysis. Such information can be additional genetic markers or topological constraints drawn from previously known evolutionary relationships. Using such strategies have, in several studies, resulted in trees comparable to those inferred from more comprehensive genetic data. It is also possible to infer community phylogenetic information using phylogenetic placement, where single barcodes are fitted within a reference tree. The reference tree can be based on another available genetic resource. These methods show great promise for the continued integration of barcode-derived phylogenetic information in community ecology. A key aspect for future development is however to adapt methods to handle datasets with a substantial number of taxa, as global datasets covering a wide range of taxa are becoming increasingly achievable with the continued development of barcoding techniques.

1. Introduction

Present-day communities are shaped by a combination of ecological, evolutionary, and stochastic processes, together determining the distribution and abundance of species (Vellend, 2010). Understanding the relative importance of these processes across environments and geographical regions is fundamental for understanding existent biodiversity and its drivers, for predicting the effects of changing environments, and for ultimately mitigating the ongoing decline of biological diversity. A key principle in community ecology is that by studying the patterns of diversity within and across communities, we can draw conclusions about which processes have been important in shaping them. Incorporating evolutionary relationships among taxa allows us to consider historical processes, providing a more comprehensive understanding of how communities have evolved over time.

The processes shaping ecological communities are often referred to as *assembly processes*, which include for example speciation, dispersal, exclusion of species due to abiotic environmental factors or biotic interactions, and neutral processes (Vellend, 2010; Weiher et al., 2011). To illustrate the influence of these forces, a widely used analogy is to compare them to filters operating at different spatial scales. The filters hierarchically sieve out species from the global species pool, which contains all existing species, to form local species communities (Figure 1). Moving from greater to smaller geographical scales, two main deterministic processes influence the remaining subset of species: environmental and biotic filtering. Environmental filtering is the result of the fundamental niche of a species, i.e. the abiotic conditions in which a species can survive and reproduce (Kraft et al., 2015). Biotic filtering is the result of interspecific interactions such as competition and facilitation, which ultimately determine the realized niche of a species (Wisz et al., 2013). In addition to these deterministic processes, stochastic processes such as ecological drift influence the final species composition in communities. Hence, species composition and abundance in local communities are the joint result of assembly processes acting across all greater scales, as well as site-specific local processes (Vellend, 2010; Weiher et al., 2011).

The effect of environmental and biotic filters depends on species traits (Mcgill et al., 2006), which are ultimately the result of their evolutionary history. Phylogenetic relationships can thus constrain species responses to different assembly processes, and phylogenetic structure constitutes and important piece of information to understand community assembly mechanisms (Pollock et al., 2012; Warren et al., 2008). In theory, environmental and biotic filtering should lead to predictable patterns of species distributions (Webb et al., 2002). Comparing observed and expected patters under various hypotheses of community assembly thus allow us to infer the relative importance of different processes. Hence, combining data of species distributions, traits, and phylogenetic relationships can enable large-scale, process-based analysis of community assembly (Münkemüller et al., 2020; Ovaskainen et al., 2017).

Recent advancements in molecular techniques have significantly reduced the costs associated with DNA analysis, resulting in a considerable increase in available genetic material (GenBank, 2024). This development can be broadly categorized into two facets: first, a transformation from phylogenetics to phylogenomics, i.e., an increase in the amount of genetic information available per specimen (Giani et al., 2020), and second, the increased application of metabarcoding, where short genetic markers are sequenced in parallel for multiple specimens (Hebert et al., 2003; Box 1). These divergent approaches present contrasting scenarios: methods that yield substantial genetic information per specimen (e.g. Kawahara et al., 2019; Misof et al., 2014) and those that produce limited information per specimen but encompass a large number of specimens (e.g. Espeland et al., 2023; Li et al., 2022).

While phylogenetic methods have advanced in parallel with the increased availability of genetic information per specimen (Dornburg et al., 2017; Yeates et al., 2016), the short sequences generated through metabarcoding poses distinct methodological challenges (Joly et al., 2014). New methodological advancements need to simultaneously address how to increase the quality of trees

inferred from limited genetic information (i.e. from short, often nearly saturated barcodes; see section 3) and develop sufficient computational performance to accommodate datasets containing a large number of taxa (Joly et al., 2014). The role of metabarcoding data has traditionally been in identification of taxa (Hebert et al., 2003), but much potential remains untapped in deriving community phylogenetic information from these versatile data. To fully harness the potential of DNA barcodes, there is an essential need to develop and refine practices in barcode-based phylogenetics (Joly et al., 2014). It is vital we understand the limitations of using barcodes in phylogenetic analyses, how to address them, and how to translate them into the realm of phylogenetic community ecology.

In this essay, I aim to provide an overview of the potential and challenges of integrating phylogenetic information derived from DNA barcodes in community ecology, focusing specifically on methods for sound phylogenetic inference. First, I will briefly outline the general use of phylogenetic information in ecological analysis. Second, I will address the key challenges of phylogenetic inference from DNA barcodes and describe how these challenges can be addressed. Finally, I will discuss the unique challenges that comes specifically from the vast size of metabarcoding data. My emphasis will be on utilizing COI barcodes for studying insects specifically, given the potential of this method for investigating this immensely diverse and largely unexplored organism group. Nonetheless, I will also incorporate instances from various other taxa and genetic repositories when appropriate.

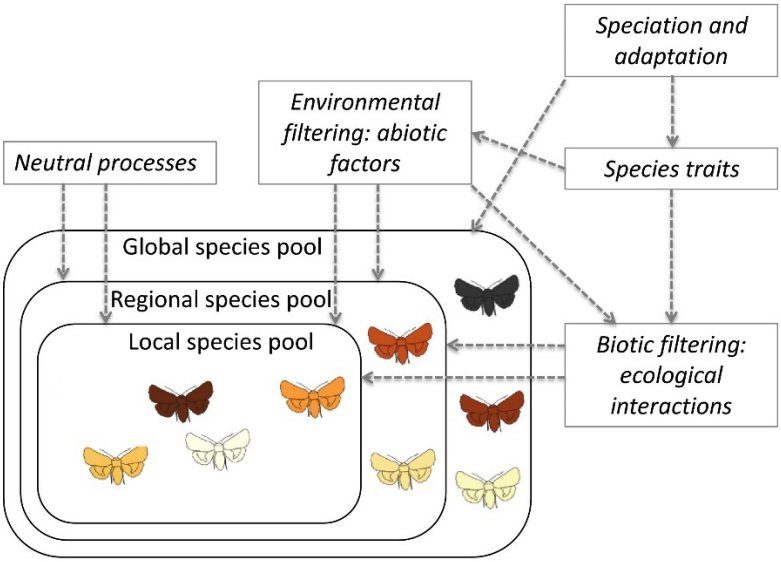


Figure 1. A conceptual figure showing the different assembly processes shaping ecological communities across varying geographical scales. The joint result of all processes is evident as the local species pool. Adapted from Ovaskainen et al. (2017).

2. Integrating phylogenetic information in ecological analysis

Phenotypic variation between all species is the result of evolution from a common ancestor. Therefore, closely related species, in general, tend to exhibit higher phenotypic similarity than distantly related species. Traits are said to show *phylogenetic signal*, i.e. 'a tendency for related species to resemble each other more than they resemble species drawn at random from the tree' (Blomberg & Garland, 2002). In a seminal paper, Webb et al. (2002) described how the phylogenetic structure of a community can indicate which processes have been particularly important for the assembly of that community. In theory, if traits follow the phylogeny, we expect a similar environment to select for similar – i.e. related

Box 1. Description of the molecular methods commonly referred to as DNA barcoding and metabarcoding.

DNA barcoding is a molecular technique for species identification using a short gene region (Hebert et al., 2003). Ideally, target regions (also called 'DNA barcodes') are chosen so that they offer substantial variation between taxa but little variation within taxa. As a result, DNA-barcodes can be uniquely identified to species by comparison with suitable reference libraries (Hebert et al., 2003). Different DNA barcodes are used for different organism groups (Kress et al., 2015). For example, COI (cytochrome c oxidase subunit I) is typically used for animals (Hebert et al., 2003; Kress et al., 2015).

The development of high-throughput sequencing, a method that allows simultaneous analysis of multiple DNA fragments, has revolutionized DNA barcoding by enabling what is known as DNA *meta*barcoding, i.e., the simultaneous sequencing of DNA barcodes for multiple taxa at once (Taberlet et al., 2012). DNA metabarcoding can be used to analyse for example environmental samples, or bulk samples containing multiple specimens (Taberlet et al., 2012). The technique enables rapid and cost-efficient analysis of diverse samples, even in cases where taxonomic expertise is lacking. DNA metabarcoding can be used for example in biodiversity assessment (e.g. Trivedi et al., 2016), or when studying species interactions (e.g. predation by analysing gut contents; Garros et al., 2008).

– species. Thus, detection of phylogenetic clustering of taxa within a local community would suggest that habitat filtering is the main structuring process (resulting in the co-occurrence of phenotypically similar species). However, there is a limit to the niche overlap of species which can coexist (MacArthur & Levins, 1967). Following the reasoning of Webb et al. (2002), strong competition between species with similar traits would result in a phylogenetically over-dispersed local community, as a result of selection favouring species with dissimilar traits – i.e. distantly related species. In short, phylogenetic relatedness is used as a proxy of ecological similarity (Mouquet et al., 2012).

Assuming correlation between phylogenetic distance and ecological affinity is, however, controversial, as many studies have failed to find evidence for this relationship (Cadotte et al., 2017). This lack of correlation could be due to different biological or methodological factors (Cadotte et al., 2017), or because some of the assumptions of the framework are not met (Gerhold et al., 2015). The assumptions and their potential issues have been described in detail by Gerhold et al. (2015). Here, I briefly mention a few of them:

1. *Competition is strongest between closely related species, and it will lead to competitive exclusion of one species.*
   This assumes that an ecological niche corresponds to a single trait state, or a single combination of trait states, and that these traits have a phylogenetic signal. However, species with different traits can utilize the same resource. If so, their phylogenetic relatedness carries little information about the strength of competition, regardless of whether the traits are phylogenetically conserved. Further, a combination of traits can sometimes be as adaptive a single trait. If these trait combinations are represented by different lineages, assembly processes are not expected to result in a phylogenetically structured community.

   It also assumes that competition always leads to exclusion. However, according to modern coexistence theory, similar species can coexist if the difference in competitive ability is small enough (Chesson, 2000). It could thus be argued that if competitive ability is trait-mediated,

and the trait shows phylogenetic signal, then we should predict that competition will lead to phylogenetic clustering, not overdispersion (Davies, 2021; Mayfield & Levine, 2010).

2. *The system is at equilibrium.*
   This assumes that community composition is stable, that is, all assembly processes have played out. In reality, both species composition and abundances will likely fluctuate over time, for example in relation to a changing climate. It further assumes that ongoing evolutionary change is irrelevant on the timescale at which we study ecological communities. That is, ecological processes are assumed to act on traits that are, in principle, fixed (Mouquet et al., 2012). However, evolutionary processes are known to affect species interactions within a few generations (Koch et al., 2014).

3. *Habitat filtering and competition are alternative processes.*
   Environmental filtering and competition can be independent from each other or operate in parallel across different spatial scales. In general, the framework is therefore only suitable for inferring the *relative* importance of the two processes.

4. *Without the influence of local processes, phylogenetic dispersion is random.*
   Other processes can cause phylogenetic dispersion to be non-random. For example, young or isolated habitats can display phylogenetically clustered communities because few species have colonized the habitat and diversification time is short. In such a case, it is not straightforward to delineate a regional species pool with which to compare the phylogenetic pattern.

Integrating phylogenetic information in ecological analyses thus call for careful consideration of appropriate geographic and temporal scales, for relevant delineation of local communities, and for matching the assumptions underlying the phylogenetic distance/ecological difference framework. However, the evolutionary history of communities is an integral part for understanding the assembly of ecological communities, and if used correctly, phylogenetic information can provide new insights about present-day biodiversity patterns (Münkemüller et al., 2020). In the remaining part of this section, I will briefly describe a few examples where researchers have used phylogenetic information to study community assembly processes.

Ramm et al. (2018) explored patterns of the stress dominance hypothesis in African squamates (snakes and lizards), using phylogenetic and functional diversity. The stress dominance hypothesis predicts that the relative influence of environmental filtering on community structuring is greater in stressful habitats, while competition is a more important process in benign environments (Weiher & Keddy, 1995). In agreement with previous findings (e.g. Cahill et al., 2008; Devictor et al., 2010), Ramm et al. (2018) found that phylogenetic diversity alone was not a useful proxy for functional diversity. Rather, it seemed to reflect mainly differences in biogeographic histories. However, patterns of functional diversity were consistent with the stress dominance hypothesis: in arid, stressful environments Ramm et al. (2018) found low functional diversity compared to in humid environments. Functional clustering in arid environments was significant also at a biome level, suggesting that a lot of the environmental filtering happens at large spatio-temporal scale. Further, relative phylogenetic diversity was high in arid environments, suggesting that the traits considered in the study (lifestyle and body size) have evolved convergently. These findings underscore the importance of multidimensional analyses and macroecological considerations in detecting general community assembly patterns (Ramm et al., 2018).

Vasar et al. (2022) used existing samples of Arbuscular Mycorrhizal fungi to study community assembly processes using taxonomic and phylogenetic structure. To infer phylogenetic structure, they built a

Neighbour-Joining tree, a relatively simple method for inferring phylogenetic relationships based on the clustering of similar taxa. They found that taxonomic community structure was primarily driven by temperature and soil pH. The same factors drove phylogenetic community structure, but the relationship was weaker, suggesting only weak niche conservatism with respect to major environmental drivers. Dispersal limitation appeared unimportant, as community structure followed the pattern of biomes rather than biogeographic realms. Similarly, Si et al. (2022) studied patterns of phylogenetic and functional diversity of mammals on islands of different area and degree of isolation. In general, island species assemblages tended to be phylogenetically clustered, and the degree of clustering increased with area and isolation. Assemblages were also found to be functionally clustered, but the pattern was weaker and displayed no relationship with area or isolation. Together, the patterns suggested that local adaptive radiation is an important process structuring communities on the islands.

Using COI barcodes in combination with other short genetic markers, Hao et al. (2020) studied the community patterns of moths in northern China. Hao et al. (2020) used a multi-locus dataset in combination with COI barcodes for robust tree inference (the method is further discussed in section 5.1), and compared patterns in communities from two separate geographic regions. From ancestral range analysis, where models of evolution are used to extrapolate back in time to infer the geographical range of the common ancestor based on current species characteristics, Hao et al. (2020) identified one region as the origin (Yanshan), from which moth communities had historically dispersed into the other region (Taihang). Hao et al. (2020) further found that the Yanshan region had a higher species diversity compared to the Taihang region, which could be the result of a period of increased diversification rate in the Yanshan region. Finally, dispersal limitation, preventing migration between Yanshan and Taihang, or the joint effect of dispersal limitation and environmental filtering, where environmental incompatibility prevent establishment of migrant species, were identified as some of the main forces structuring the current moth communities in the two regions. Overall, these authors concluded that an analysis considering both historical and contemporary factors is useful for identifying factors shaping communities.

3. Potential and limitations of the COI barcode for phylogenetic inference

As previously described, lower costs and increased performance of DNA barcoding and metabarcoding have resulted in an upsurge of available genetic data from ecological communities. Utilizing DNA barcode not only for species identification, but also for inferring evolutionary relationships between species in communities, can help increase our understanding of community assembly processes and biodiversity drivers.

The most common genetic marker for DNA barcoding of animals is a part of the mitochondrial gene coding for the enzyme cytochrome c oxidase I (COI) (Hebert et al., 2003). A fundamental criterion for a suitable barcode for species identification is that the degree of interspecific variation is substantially greater than intraspecific variation. This has been shown for COI within several taxonomic groups (Hebert et al., 2003). In addition, the COI barcode region has several other characteristics that makes it suitable:

- *ubiquity*: the COI gene can be found in many lineages, even across taxonomic domains,
- *conserved and variable regions*: the gene contains enough variation to separate between (most) species, while still having conserved regions that facilitate alignment across higher taxonomic groups,
- *few insertions and deletions*: which makes alignment relatively straightforward,
- *part of the mitochondrial genome*: uniparental inheritance decreases variation due to a lack of recombination, and multiple gene copies per cell facilitate amplification,

Box 2. Description of common metrics to measure the phylogenetic structure of communities

Ecologists tend to ask three types of questions regarding phylogenetic relationships within or between communities: questions about richness, divergence, or regularity (Tucker et al., 2017). Metrics of richness quantifies the phylogenetic differences present in a sample, divergence the average phylogenetic distance between taxa, and regularity how evenly distanced taxa are from each other across the tree (Figure 2). There are multiple measures for each of these dimensions, but Tucker et al. (2017) identified three measures that align most closely with each dimension; phylogenetic diversity to measure richness, mean pairwise distance to measure divergence, and variation of pairwise distances to measure regularity.

Two popular metrics used to quantify the phylogenetic distribution of taxa in a sample relative to a selected species pool are the net relatedness index (NRI) and nearest taxon index (NTI) (Webb, 2000; Webb et al., 2002). NRI quantifies overall clustering of taxa in a tree, by measuring the mean pairwise phylogenetic distance of taxa in a sample. NTI, on the other hand, estimates the degree of terminal clustering, independent from the degree of clustering deeper in the tree, by measuring the phylogenetic distance to the nearest taxon for each taxon in the sample. A common practice is to compare observed values of NRI and NTI to a null model, to determine if the phylogenetic structure is different than what is expected by chance. Higher values of NRI and/or NTI signifies a higher degree of clustering than expected, while negative values suggest overdispersion.

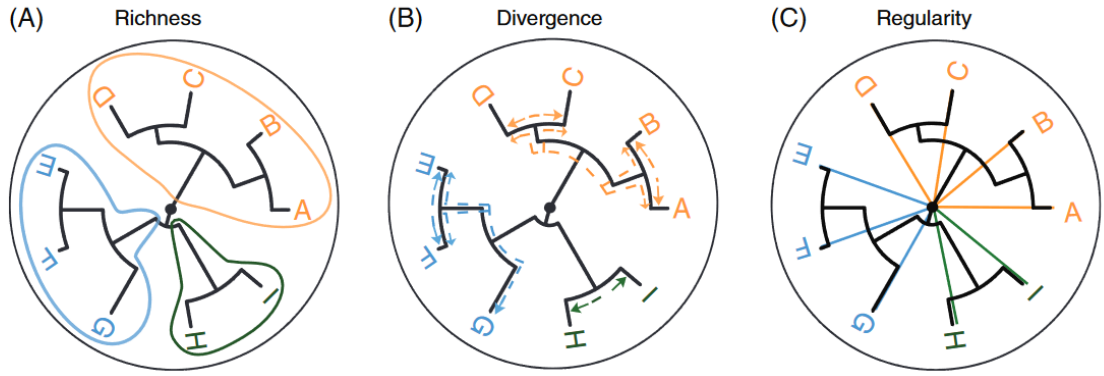


Figure 2. Conceptual figure of the three dimensions of phylogenetic structure in a tree: (a) richness, (b) divergence, and (c) regularity. Each dimension is associated with a plethora of different metrics, but most closely aligned are phylogenetic diversity, mean pairwise distance, and variation of pairwise distances, respectively.

- *reference databases*: after two decades of use, reference databases have particularly great taxa coverage for COI.

However, using DNA barcoding data for phylogenetic analysis brings with it a set of limitations. First, barcodes are generally very short sequences. The full-length COI barcode is 658 bp long (Hebert et al., 2003), and only a subset of that is amplified in metabarcoding, to match the length compatible with current high-throughput sequencing platforms (Porter & Hajibabaei, 2020). It is clear that the short sequences generated by (meta-)barcoding contain limited phylogenetic information compared to the large multiple sequence alignments used for determining relationships between for example insect families (2.2 million bp for Lepidoptera (Kawahara et al., 2019), 7 000 - 30 000 bp for Diptera; (Wiegmann et al., 2011)). However, the precise limitation of how old divergences can be recovered from barcodes are yet to be determined. Some studies suggest that the COI barcode region is sufficient for accurately recovering monophyletic groups below family level (Boyle & Adamowicz, 2015; Wilson,

2011), or of divergences occurring less than 3-4 million years ago (Trunz et al., 2016). Second, further adding to this challenge, barcodes are chosen based on having a mutation rate optimal for species identification, and they quickly become too saturated to study old divergences. In addition, strong functional constraints on the amino acid level have been shown to cause parallel or convergent evolution between taxa (Pentinsaari et al., 2016), further limiting the variation of the short sequence. Thus, the utility of the COI barcode when using traditional phylogenetic methods is generally limited to recent divergences. These challenges can however be addressed by using alternative methods for building phylogenies, which I describe further in section 5.1.

4. Evaluating phylogenetic accuracy

During the advancement of phylogenetic methods, evaluating the accuracy and robustness of inferred tree is, of course, pivotal. However, as we can never know the true evolutionary relationship between taxa (except in the case of simulated data), tree evaluation is a non-trivial task. In some cases, we can make relatively conservative assumptions about which trees are more likely to reflect true relationships. For example, Talavera et al. (2022) inferred trees either from COI barcodes alone for all taxa, or adding multiple genetic markers for a subset of taxa to aid inference of deep phylogenetic relationships. In this case, when comparing the resulting trees, the authors assumed that the multilocus information increased the tree accuracy, which is likely to be true. However, in many cases, finding a "ground truth" to compare trees against is not straightforward. Therefore, alternative approaches or proxy measures are often used to evaluate phylogenies.

One aspect of evaluating phylogenies is determining the sensitivity of the resulting tree to the choice of the specific method. This can be done by generating multiple trees using for example different models of evolution or different methods for tree inference. If different methods recover similar trees, one can infer that the method choice does not have a large impact on the final result. Should the trees differ largely between methods, there is a need to decipher the mechanisms behind this and evaluate more carefully which method to use. When presenting phylogenetic trees, it is essential to showcase which components of the tree, e.g. splits in the topology or age of nodes, are more or less uncertain. Some methods inherently offer these measures, such as the probability distributions generated by Bayesian methods. In other instances, estimating uncertainty is possible using methods like bootstrapping or jack-knifing. With these approaches, multiple trees are inferred by resampling or using subsets of the data. The frequency of occurrence of each branch within the set of inferred trees can serve as a measure of support for those branches.

Finally, one can estimate the congruence between trees produced using different data sources, to evaluate the sensitivity of the result to the choice of data. However, direct comparison of topological congruence between trees is strongly affected by the number of taxa; as the number of taxa increases, the probability of a group of taxa forming a clade decreases. Instead, the Taxon Consistency Index (TCI) and Taxon Retention Index (TRI) have been proposed as measures of taxonomic congruence (Figure 3; Wilson, 2011). Both measures how selected groups of taxa (e.g. taxonomic groups such as families or genera) cluster in the tree, but only TRI scales with number of taxa.

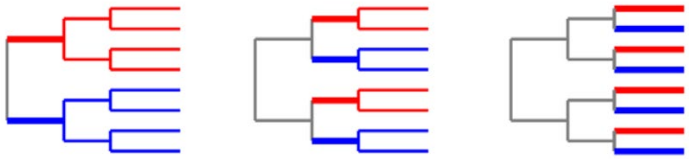
Taxon consistency index TCI = M^t/S^t

M^t = # of concordance groups in the test (i.e. # of families)

S^t = minimum # of clades exhibited by concordance groups on cladogram

Taxon retention index TRI = $(G^t-S^t)/(G^t-M^t)$

G^t = # species in datamatrix



TCI	2/2 = **1**	2/4 = **0.5**	2/8 = **0.25**
TRI	(8-2)/(8-2) = **1**	(8-4)/(8-2)= **0.66**	(8-8)/(8-2) = **0**

Figure 3. Taxon consistency index (TCI) and taxon retention index (TRI) both measure how well expected clades cluster in a phylogenetic tree. Figure from Wilson (2011).

5. Using DNA barcodes in phylogenetic community analysis

Despite the limited phylogenetic signal in DNA barcodes, many studies have used barcode data from ecological communities for phylogenetic analyses. Two main methods are used: *de novo* phylogenetic inference, where barcodes are used to infer phylogenetic trees, and phylogenetic placement, where single barcodes are placed within a reference tree. Here, I will discuss different methodological adaptations to utilize these two approaches for DNA barcode data. Further, I will outline some of the future challenges of using DNA barcodes in community phylogenetics, including some of the implications of increasingly large datasets.

5.1 Phylogenetic inference *de novo*

As previously described, DNA barcodes are chosen for their utility for species identification, and thus the phylogenetic signal for most barcodes is strongest for recent divergences (e.g. Boyle & Adamowicz, 2015; Trunz et al., 2016; Wilson, 2011). For phylogenetic inference, the limitation in the signal means that DNA barcodes are mainly useful for resolving relationships between species and genera, while inferring deeper relationships such as between families or orders becomes increasingly difficult. In most cases, DNA barcodes needs to be complemented with additional information to resolve deep evolutionary relationships. However, some evidence suggests that the sheer abundance of DNA barcodes can facilitate phylogenetic inference, by giving a more complete evolutionary record than what is available with sparse taxon sampling. Dense taxon sampling could break up long branches, thus resulting in more accurate trees. Here, I will describe two strategies for complementing DNA barcodes with additional information: (1) using topological constraints, or a so-called *backbone tree,* to constrain the tree search, and (2) using multiple genetic markers. Further, I will briefly discuss the effects of dense taxon sampling on phylogenetic inference.

Topological constraints limit or direct the branching pattern of a phylogenetic tree, thus making it possible to take into account previously known relationships between taxa. Constraints can be defined using different strategies, such as using taxonomic groups to define monophyletic groups or deriving information from other phylogenetic trees containing some of the query taxa. The latter is often referred to as a 'backbone tree' and provides a framework where additional taxa can be added to the tree without altering the deeper relationships in the tree. Chesters (2020) used distinct genetic datasets to define topological constraints in three hierarchical steps when building an insect phylogeny containing more than 69 000 species. The deepest part of the phylogeny, including for example

relationships between insect orders, was inferred from transcriptomes. Thereafter, the tree was expanded by adding mitogenomes and DNA barcodes in two sequential stages of phylogenetic inference, each constrained by the backbone tree created in the previous stage. As the three stages was based on genetic resources that decreased in phylogenetic information but increased in taxon richness, the method allowed the addition of a large number of taxa while maintaining the topology from more phylogenetically robust datasets. To optimize taxa coverage, nine different loci were used in the final tree layer, including the COI barcode and 16S ribosomal RNA. However, due to the sparsity of available data from the nine loci (8.9% of all loci and taxa), the final matrix contained only an average of <770 base pairs (bp) per species, which is not much more than the standard COI barcode region (658 bp). The combination of genetic markers can still provide more phylogenetic power than the single COI barcode, as a result of varying evolutionary rate, which I describe later in this section.

Boyle and Adamowicz (2015) wanted to evaluate the effect of phylogenetic tree construction method on commonly used community metrics of phylogenetic diversity and structure (NRI and NTI; see Box 2). They used a dataset of larval Trichopterans from 46 different species and inferred Bayesian phylogenetic trees using either multi-locus data, COI only, or COI and a family-level backbone tree derived from another data source. For COI only trees, they also inferred a tree using Neighbour Joining. Using the multi-locus tree as ground truth, they found that the use of a family-level backbone tree resulted in the best estimation of phylogenetic community structure. Neighbour Joining resulted in the worst congruence, while Bayesian trees base on COI only performed surprisingly well.

One of the determinants of how old divergences can be recovered with a specific genetic sequence is the *evolutionary rate* of the sequence. Fast evolving sequences quickly becomes saturated, rendering them useless for inferring old divergences (e.g. Boyle & Adamowicz, 2015; Wilson, 2011). At the other extreme, slowly evolving sequences can display insufficient variation to separate between closely related species. In phylogenetic analysis, using multiple genetic markers with varying evolutionary rate can thus enrich the available information, enhancing the accuracy of inferred trees. Talavera et al. (2022) demonstrated this for the subtribe *Polyommatina* (*Lepidoptera: Lycaenidae*), using a dataset with complete sampling (i.e. of all taxa) of the COI barcode complemented with incomplete sampling of eight additional markers. The incomplete sampling of multiple loci included markers such as cytochrome oxidase II (COII), 28S ribosome unit (28S), and internal transcribed spacer 2 (ITS2). Using a phylogeny derived from only the taxa with multilocus data (i.e. 8% of the total count of taxa) as a reference, they found that genus level node recovery improved when inferring phylogenies from the full dataset (COI barcodes + multilocus data) than when using COI barcodes alone. Further, interspecies relationships changed when adding the multiocus data, suggesting that the additional information resulted in meaningful differences also within genera.

To find what proportion of taxa need to be included in the multilocus dataset to improve phylogenetic inference, Talavera et al. (2022) simulated data of gene evolution, using parameters based on observed rates of evolution within *Polyommatina*. They simulated datasets with cover of the multilocus data varying between 0-100%, thus representing datasets ranging from COI barcodes only to complete sampling of all loci. Missing data was either randomly distributed across taxa or non-random across genera, to simulate strategically sampled data. As expected, they found that the proportion of correctly resolved nodes and branch lengths improved with increasing cover of the multilocus data. The relationship was however not linear, rather there was a steep improvement when multilocus data cover was around 5-10% and sampling was strategic. Additional data beyond this level resulted in minimal improvement in accuracy, and trees with 5-10% of strategically sampled multilocus data was comparable with trees from a matrix with 100% data cover. In contrast, trees without any data of additional markers were of considerably lower quality, failing to recover accurate relationships between genera. To guide the strategic sampling, one could use DNA barcodes to identify particularly divergent lineages which would benefit from additional information (Trunz et al., 2016), thus iteratively increasing the power of the data.

In alignment with the preceding results, including an additional nuclear marker to complement the COI barcode increased the support for subgenera clade recovery within the tribe *Megachilini* (*Hymenoptera: Megachilidae*) (Trunz et al., 2016). They estimated that the COI barcode contained phylogenetic signal limited to divergences 3-4 million years ago, while adding a nuclear marker (LW-rhodopsin) increased the signal to divergences around 15 million years ago. Interestingly, when inferring trees from COI barcodes alone, they found that Bayesian methods sometimes recovered monophyletic groups of divergences older than 4 million years in agreement with current taxonomy. However, the method also recovered clusters contradicting current taxonomic knowledge with similar posterior probability. This could suggest either a disagreement between current taxonomy and phylogenetic relationships between species, or a methodological artifact creating false confidence in an erroneous tree topology.

In many ecological studies drawing on evolutionary information, phylogenetic trees are mainly used to calculate different metrics of phylogenetic diversity of communities. In such cases, the importance of tree quality thus mainly depends on the effect of tree accuracy on such metrics. Liu et al. (2019) evaluated the effect of using multiple genetic markers (rbcL, matK, ITS, and ITS2) or a family-level backbone tree to improve tree inference on phylogenetic diversity metrics of plant communities. They found that using multiple markers, specifically one conserved and one variable, increased the robustness of the resulting phylogeny. Further, the use of topological constraints drawn from a family-level backbone tree affected diversity metrics based on topology. However, the direction of the effect varied between metrics, highlighting the need for evaluating the sensitivity of the metrics with respect to method choice.

One hypothesis state that dense taxon sampling can facilitate phylogenetic inference by breaking up long branches in the tree. This is a compelling proposition that could reinforce the utility of DNA barcodes for phylogenetic inference. If true, the sheer volume of available barcodes could facilitate accurate phylogenetic inference. Wilson (2011) tested whether increased taxon sampling improved the quality of a phylogeny inferred from COI DNA barcodes only, using families of Macrolepidoptera as a case study. However, even with dense sampling (500 species per family), Wilson (2011) was unable to recover monophyly within families. Below family level, taxon clusters remained similar when taxon sampling increased, suggesting that the COI barcode has a stronger phylogenetic signal below family level. The results also suggested that increased taxon sampling did not break up long branches, as average genetic distance between taxa remained unchanged when sampling increased.

5.2 Phylogenetic placement

Phylogenetic placement, also known as evolutionary placement, refers to a set of methods that involve fitting a DNA sequence – the query sequence – within a reference tree derived from a separate data source (Czech et al., 2022). The method is primarily used to render a taxonomic classification of sequences and can be especially useful in cases where close relatives are not present in reference databases. However, phylogenetic placement also provides information about the evolutionary relationship between the query and reference sequences, much like a phylogenetic tree. When comparing phylogenetic placement with local *de novo* tree inference with regards to taxonomic classification across the class Insecta, Chesters (2017) found that phylogenetic placement produced more accurate classifications. In addition to the improvement in accuracy, phylogenetic placement is generally much faster than *de novo* phylogenetic tree inference, since none of the query sequences are analysed against each other. The approach of analysing each query sequence independent from the others make the method easily parallelizable, i.e. it can run on multiple computer cores simultaneously, thus significantly speeding up the analysis. Nevertheless, this implies that the method does not provide any information about the evolutionary relationship between two query sequences. If many query sequences belong to closely related taxa it can thus be difficult to determine how they

relate to each other, unless the reference tree is densely sampled, i.e. contain many reference sequences within that specific taxonomic group.

Assessing the feasibility of using phylogenetic placement in community ecology studies, using COI barcodes from Coleoptera, Li et al. (2022) compared the use of *de novo* phylogenetic trees to phylogenetic placement on a robust reference tree. The authors investigated how community clustering across latitude and season changed depending on whether phylogenetic relationships were included or not, and found that including the evolutionary relationships between taxa changed the predominant axis of variation among communities. That is, without including phylogenetic information, samples clustered based on latitude (i.e. communities in the north were more similar to other northern communities than to southern communities) and with phylogenetic information communities clustered based on season (i.e. summer samples were more similar to other summer samples than winter samples). However, the result was robust to the choice of method between *de novo* tree inference and phylogenetic placement.

Other studies have also successfully used phylogenetic placement algorithms to study phylogenetic patterns in communities. Wang et al. (2020, 2022) studied community structure of Lepidopteran caterpillars on different tree species. Using phylogenetic placement to infer patterns of evolutionary relationships, they found that the community composition of caterpillars was affected by the phylogenetic host composition and related functional traits, such as defences and palatability (Wang et al., 2020). This pattern could be the result of co-evolution of Lepidopteran herbivores and their host plants. They further found a negative correlation between co-occurrence probability and phylogenetic distance of both caterpillars and plants, suggesting that closely related species occur together more often than distantly related species, and that they prefer plants of similar evolutionary history (Wang et al., 2022). This implies the role of environmental filtering as a structuring force in the caterpillar community, and further suggest co-evolution between certain herbivore and plant traits.

5.3 Future avenues for development

Technological development of metabarcoding and reduced sequencing costs have transformed DNA-based studies of community ecology, enabling processing of higher number of samples and massive amounts of genetic data. While this offers unprecedented opportunities for phylogenetic analysis, it comes with the challenge of developing phylogenetic methods to simultaneously increase the quality of tree inference from short DNA sequences and increasing the amount of data that can be handled in a single analysis.

Insect phylogenies based on DNA barcodes are currently expanding to cover a larger number of taxa. For example, Espeland et al. (2023) built a phylogeny for 1280 specimen representing 449 species of the subtribe *Euptychiina* (*Lepidoptera: Satyrini*) and 39 outgroups. However, inferring phylogenies from COI barcodes within lower-rank taxonomic groups, such as subtribes, is generally easier than inferring relationships within for example families or orders, since divergences within subtribes are generally relatively recent. Covering a wider taxonomic range, Tiusanen et al. (2019) used COI barcodes to study flower-visitor communities in arctic environments. Beyond resolving species richness of communities, the authors inferred a phylogenetic tree for 1314 species across more than six insect orders. Based on the recommendations of Boyle and Adamowicz (2015), Tiusanen et al. (2019) used a backbone tree to constrain deep evolutionary relationships. Based on this phylogeny, the authors calculated phylogenetic diversity of communities spread across Arctic areas, and found for example that patterns of phylogenetic history is in concordance with the geological history of the region.

Li et al. (2022) used DNA barcodes from Coleoptera specimen caught in Malaise traps to infer phylogenetic relationships using both phylogenetic placement and de novo tree inference. Li et al. (2022) did this for each unique sequence in their samples, amounting to about 7 000 sequences. To allow analysis of such a large dataset, they used Maximum Likelihood methods. Maximum Likelihood

is generally more computationally efficient than for example Bayesian methods, generally making it the most feasible choice for inferring large phylogenies.

The most comprehensive insect taxonomies, to my knowledge, also use DNA barcodes as the final data tier to increase taxon sampling across the tree (Chesters, 2017, 2020). As described earlier, the tree was built using three tiers of genetic data; transcriptomes, mitogenomes, and DNA barcodes, each tier constraining the topology of the tree inferred from the subsequent tier. In the first instance of the Insecta tree, it includes more than 49 000 species across 771 families (Chesters, 2017). In the second iteration, the species count was increased to more than 69 000 species (Chesters, 2020). However, due to difficulties taxonomically placing species-level data, the higher-rank taxonomic cover decreased from 771 to 503 families.

Scaling up analyses to such extensive datasets often demands prioritizing speed over accuracy. Though methods such as Maximum Likelihood perform well even during fast analyses, understanding the limitations and implications of this trade-off is crucial. Important questions that remain underexplored include for example how the choice of phylogenetic method affects the utility of COI barcodes in tree inference. Are more sophisticated, computationally demanding methods better at recovering evolutionary relationships between taxa based on limited phylogenetic information, or are they creating artificial confidence in erroneous topologies? Further, how does differences between methods for phylogenetic inference translate into subsequent ecological analyses? Identifying cases where method choice might impact end results and conclusions is essential for advancing the use of DNA barcodes in phylogenetic community ecology.

Importantly, the development of large-scale phylogenies covering a broad taxonomic range might facilitate the use of phylogenies in studies of local communities. Trees such as Chesters' comprehensive *Insecta* phylogeny (2017, 2020), or the Time Tree of Life (Kumar et al., 2022), are valuable resources that can aid tree inference for focal taxa by providing a robust backbone. In this regard, a crucial challenge lies in determining the extent of phylogenetic information retrievable from DNA barcodes. While some evidence suggest that genus-level backbone trees are necessary for accurate phylogenetic inference (Boyle & Adamowicz, 2015; Trunz et al., 2016; Wilson, 2011), there is likely variation across taxonomic groups. Should a genus-level tree be necessary, that poses a significant hurdle for employing backbone-constrained inference in areas with sparse data coverage, such as the tropics. Generally, genetic data availability is heavily skewed towards Europe and North America (Chesters, 2020).

6. Conclusion

The popularity and widespread use of COI barcodes for species identification has contributed to building one of the most taxonomically diverse genetic repositories to date. Harnessing this genetic resource for phylogenetic analysis presents exciting opportunities for community ecology that were previously unattainable. In this essay, I have shown that despite substantial challenges for phylogenetic inference from barcodes, imposed mainly by a limited phylogenetic signal, method development has enabled robust inference of phylogenetic relationships within communities. To fully utilize the ever-expanding resource that is COI barcodes, it is imperative to continue investigating the limitations of phylogenetic inference based on barcodes across a wide range of taxa. In parallel, the continued developing of methods for augmenting phylogenetic inference from limited genetic material should aim for adapting methods to handle datasets with a large number of taxa, as such datasets are becoming increasingly common. To date, numerous studies have made use of COI barcodes to infer phylogenetic community patterns. Future development of this research area holds great potential to further facilitate large-scale community ecology studies, opening new horizons for ecological research.

7. References

Boyle, E. E., & Adamowicz, S. J. (2015). Community Phylogenetics: Assessing Tree Reconstruction Methods and the Utility of DNA Barcodes. *PLOS ONE*, *10*(6), e0126662. https://doi.org/10.1371/journal.pone.0126662

Cadotte, M. W., Davies, T. J., & Peres-Neto, P. R. (2017). Why phylogenies do not always predict ecological differences. *Ecological Monographs*, *87*(4), 535–551. https://doi.org/10.1002/ecm.1267

Cahill, J. F., Kembel, S. W., Lamb, E. G., & Keddy, P. A. (2008). Does phylogenetic relatedness influence the strength of competition among vascular plants? *Perspectives in Plant Ecology, Evolution and Systematics*, *10*(1), 41–50. https://doi.org/10.1016/j.ppees.2007.10.001

Chesson, P. (2000). Mechanisms of Maintenance of Species Diversity. *Annual Review of Ecology and Systematics*, *31*(1), 343–366. https://doi.org/10.1146/annurev.ecolsys.31.1.343

Chesters, D. (2017). Construction of a Species-Level Tree of Life for the Insects and Utility in Taxonomic Profiling. *Systematic Biology*, syw099. https://doi.org/10.1093/sysbio/syw099

Chesters, D. (2020). The phylogeny of insects in the data-driven era. *Systematic Entomology*, *45*(3), 540–551. https://doi.org/10.1111/syen.12414

Czech, L., Stamatakis, A., Dunthorn, M., & Barbera, P. (2022). Metagenomic Analysis Using Phylogenetic Placement—A Review of the First Decade. *Frontiers in Bioinformatics*, *2*, 871393. https://doi.org/10.3389/fbinf.2022.871393

Davies, T. J. (2021). Ecophylogenetics redux. *Ecology Letters*, *24*(5), 1073–1088. https://doi.org/10.1111/ele.13682

Devictor, V., Mouillot, D., Meynard, C., Jiguet, F., Thuiller, W., & Mouquet, N. (2010). Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: The need for integrative conservation strategies in a changing world. *Ecology Letters*, *13*(8), 1030–1040. https://doi.org/10.1111/j.1461-0248.2010.01493.x

Dornburg, A., Townsend, J. P., & Wang, Z. (2017). Maximizing Power in Phylogenetics and Phylogenomics: A Perspective Illuminated by Fungal Big Data. In *Advances in Genetics* (Vol. 100, pp. 1–47). Elsevier. https://doi.org/10.1016/bs.adgen.2017.09.007

Espeland, M., Nakahara, S., Zacca, T., Barbosa, E. P., Huertas, B., Marín, M. A., Lamas, G., Benmesbah, M., Brévignon, C., Casagrande, M. M., Fåhraeus, C., Grishin, N., Kawahara, A. Y., Mielke, O. H. H., Miller, J. Y., Nakamura, I., Navas, V., Patrusky, B., Pyrcz, T. W., … Willmott, K. R. (2023). Combining target enrichment and Sanger sequencing data to clarify the systematics of the diverse Neotropical butterfly subtribe Euptychiina (Nymphalidae, Satyrinae). *Systematic Entomology*, *48*(4), 498–570. https://doi.org/10.1111/syen.12590

Garros, C., Ngugi, N., Githeko, A. E., Tuno, N., & Yan, G. (2008). Gut content identification of larvae of the *Anopheles gambiae* complex in western Kenya using a barcoding approach. *Molecular Ecology Resources*, *8*(3), 512–518. https://doi.org/10.1111/j.1471-8286.2007.02013.x

GenBank. (2024, January 19). *GenBank and WGS Statistics*. https://www.ncbi.nlm.nih.gov/genbank/statistics/

Gerhold, P., Cahill, J. F., Winter, M., Bartish, I. V., & Prinzing, A. (2015). Phylogenetic patterns are not proxies of community assembly mechanisms (they are far better). *Functional Ecology*, *29*(5), 600–614. https://doi.org/10.1111/1365-2435.12425

Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*, *18*, 9–19. https://doi.org/10.1016/j.csbj.2019.11.002

Hao, M., Jin, Q., Meng, G., Yang, C., Yang, S., Shi, Z., Tang, M., Liu, S., Li, Y., Zhang, D., Su, X., Shih, C., Sun, Y., Zhou, X., & Zhang, A. (2020). Regional assemblages shaped by historical and contemporary factors: Evidence from a species-rich insect group. *Molecular Ecology*, *29*(13), 2492–2510. https://doi.org/10.1111/mec.15412

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Joly, S., Davies, T. J., Archambault, A., Bruneau, A., Derry, A., Kembel, S. W., Peres-Neto, P., Vamosi, J., & Wheeler, T. A. (2014). Ecology in the age of DNA barcoding: The resource, the promise and the challenges ahead. *Molecular Ecology Resources*, *14*(2), 221–232. https://doi.org/10.1111/1755-0998.12173

Kawahara, A. Y., Plotkin, D., Espeland, M., Meusemann, K., Toussaint, E. F. A., Donath, A., Gimnich, F., Frandsen, P. B., Zwick, A., dos Reis, M., Barber, J. R., Peters, R. S., Liu, S., Zhou, X., Mayer, C., Podsiadlowski, L., Storer, C., Yack, J. E., Misof, B., & Breinholt, J. W. (2019). Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proceedings of the National Academy of Sciences*, *116*(45), 22657–22663. https://doi.org/10.1073/pnas.1907847116

Koch, H., Frickel, J., Valiadi, M., & Becks, L. (2014). Why rapid, adaptive evolution matters for community dynamics. *Frontiers in Ecology and Evolution*, *2*. https://doi.org/10.3389/fevo.2014.00017

Kraft, N. J. B., Adler, P. B., Godoy, O., James, E. C., Fuller, S., & Levine, J. M. (2015). Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology*, *29*(5), 592–599. https://doi.org/10.1111/1365-2435.12345

Kress, W. J., García-Robledo, C., Uriarte, M., & Erickson, D. L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology & Evolution*, *30*(1), 25–35. https://doi.org/10.1016/j.tree.2014.10.008

Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., Stecher, G., & Hedges, S. B. (2022). TimeTree 5: An Expanded Resource for Species Divergence Times. *Molecular Biology and Evolution*, *39*(8), msac174. https://doi.org/10.1093/molbev/msac174

Li, Z., Linard, B., Vogler, A. P., Yu, D. W., & Wang, Z. (2022). Phylogenetic diversity only weakly mitigates climate-change-driven biodiversity loss in insect communities. *Molecular Ecology*, mec.16747. https://doi.org/10.1111/mec.16747

Liu, J., Liu, J., Shan, Y.-X., Ge, X., & Burgess, K. S. (2019). The use of DNA barcodes to estimate phylogenetic diversity in forest communities of southern China. *Ecology and Evolution*, *9*(9), 5372–5379. https://doi.org/10.1002/ece3.5128

MacArthur, R., & Levins, R. (1967). The Limiting Similarity, Convergence, and Divergence of Coexisting Species. *The American Naturalist*, *101*(921), 377–385. https://doi.org/10.1086/282505

Mayfield, M. M., & Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities: Phylogeny and coexistence. *Ecology Letters*, *13*(9), 1085–1093. https://doi.org/10.1111/j.1461-0248.2010.01509.x

Mcgill, B., Enquist, B., Weiher, E., & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution*, *21*(4), 178–185. https://doi.org/10.1016/j.tree.2006.02.002

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., Frandsen, P. B., Ware, J., Flouri, T., Beutel, R. G., Niehuis, O., Petersen, M., Izquierdo-Carrasco, F., Wappler, T., Rust, J., Aberer, A. J., Aspöck, U., Aspöck, H., Bartel, D., … Zhou, X. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science*, *346*(6210), 763–767. https://doi.org/10.1126/science.1257570

Mouquet, N., Devictor, V., Meynard, C. N., Munoz, F., Bersier, L.-F., Chave, J., Couteron, P., Dalecky, A., Fontaine, C., Gravel, D., Hardy, O. J., Jabot, F., Lavergne, S., Leibold, M., Mouillot, D., Münkemüller, T., Pavoine, S., Prinzing, A., Rodrigues, A. S. L., … Thuiller, W. (2012). Ecophylogenetics: Advances and perspectives. *Biological Reviews*, *87*(4), 769–785. https://doi.org/10.1111/j.1469-185X.2012.00224.x

Münkemüller, T., Gallien, L., Pollock, L. J., Barros, C., Carboni, M., Chalmandrier, L., Mazel, F., Mokany, K., Roquet, C., Smyčka, J., Talluto, M. V., & Thuiller, W. (2020). Dos and don'ts when inferring assembly rules from diversity patterns. *Global Ecology and Biogeography*, *29*(7), 1212–1229. https://doi.org/10.1111/geb.13098

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., & Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, *20*(5), 561–576. https://doi.org/10.1111/ele.12757

Pentinsaari, M., Salmela, H., Mutanen, M., & Roslin, T. (2016). Molecular evolution of a widely-adopted taxonomic marker (COI) across the animal tree of life. *Scientific Reports*, *6*(1), 35275. https://doi.org/10.1038/srep35275

Pollock, L. J., Morris, W. K., & Vesk, P. A. (2012). The role of functional traits in species distributions revealed through a hierarchical model. *Ecography*, *35*(8), 716–725. https://doi.org/10.1111/j.1600-0587.2011.07085.x

Porter, T. M., & Hajibabaei, M. (2020). Putting COI Metabarcoding in Context: The Utility of Exact Sequence Variants (ESVs) in Biodiversity Analysis. *Frontiers in Ecology and Evolution*, *8*, 248. https://doi.org/10.3389/fevo.2020.00248

Ramm, T., Cantalapiedra, J. L., Wagner, P., Penner, J., Rödel, M.-O., & Müller, J. (2018). Divergent trends in functional and phylogenetic structure in reptile communities across Africa. *Nature Communications*, *9*(1), 4697. https://doi.org/10.1038/s41467-018-07107-y

Si, X., Cadotte, M. W., Davies, T. J., Antonelli, A., Ding, P., Svenning, J., & Faurby, S. (2022). Phylogenetic and functional clustering illustrate the roles of adaptive radiation and dispersal filtering in jointly shaping late-Quaternary mammal assemblages on oceanic islands. *Ecology Letters*, *25*(5), 1250–1262. https://doi.org/10.1111/ele.13997

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. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Talavera, G., Lukhtanov, V., Pierce, N. E., & Vila, R. (2022). DNA Barcodes Combined with Multilocus Data of Representative Taxa Can Generate Reliable Higher-Level Phylogenies. *Systematic Biology*, *71*(2), 382–395. https://doi.org/10.1093/sysbio/syab038

Tiusanen, M., Huotari, T., Hebert, P. D. N., Andersson, T., Asmus, A., Bêty, J., Davis, E., Gale, J., Hardwick, B., Hik, D., Körner, C., Lanctot, R. B., Loonen, M. J. J. E., Partanen, R., Reischke, K., Saalfeld, S. T., Senez-Gagnon, F., Smith, P. A., Šulavík, J., … Roslin, T. (2019). Flower-visitor communities of an arcto-alpine plant—Global patterns in species richness, phylogenetic diversity and ecological functioning. *Molecular Ecology*, *28*(2), 318–335. https://doi.org/10.1111/mec.14932

Trivedi, S., Aloufi, A. A., Ansari, A. A., & Ghosh, S. K. (2016). Role of DNA barcoding in marine biodiversity assessment and conservation: An update. *Saudi Journal of Biological Sciences*, *23*(2), 161–171. https://doi.org/10.1016/j.sjbs.2015.01.001

Trunz, V., Packer, L., Vieu, J., Arrigo, N., & Praz, C. J. (2016). Comprehensive phylogeny, biogeography and new classification of the diverse bee tribe Megachilini: Can we use DNA barcodes in phylogenies of large genera? *Molecular Phylogenetics and Evolution*, *103*, 245–259. https://doi.org/10.1016/j.ympev.2016.07.004

Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., Grenyer, R., Helmus, M. R., Jin, L. S., Mooers, A. O., Pavoine, S., Purschke, O., Redding, D. W., Rosauer, D. F., Winter, M., & Mazel, F. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology: A guide to phylogenetic metrics for ecology. *Biological Reviews*, *92*(2), 698–715. https://doi.org/10.1111/brv.12252

Vasar, M., Davison, J., Sepp, S.-K., Oja, J., Al-Quraishy, S., Bueno, C. G., Cantero, J. J., Fabiano, E. C., Decocq, G., Fraser, L., Hiiesalu, I., Hozzein, W. N., Koorem, K., Moora, M., Mucina, L., Onipchenko, V., Öpik, M., Pärtel, M., Phosri, C., … Zobel, M. (2022). Global taxonomic and phylogenetic assembly of AM fungi. *Mycorrhiza*, *32*(2), 135–144. https://doi.org/10.1007/s00572-022-01072-7

Vellend, M. (2010). Conceptual Synthesis in Community Ecology. *The Quarterly Review of Biology*, *85*(2), 183–206. https://doi.org/10.1086/652373

Wang, M., Li, Y., Chesters, D., Bruelheide, H., Ma, K., Guo, P., Zhou, Q., Staab, M., Zhu, C., & Schuldt, A. (2020). Host functional and phylogenetic composition rather than host diversity structure

plant–herbivore networks. *Molecular Ecology*, *29*(14), 2747–2762. https://doi.org/10.1111/mec.15518

Wang, M., Yan, C., Luo, A., Li, Y., Chesters, D., Qiao, H., Chen, J., Zhou, Q., Ma, K., Bruelheide, H., Schuldt, A., Zhang, Z., & Zhu, C. (2022). Phylogenetic relatedness, functional traits, and spatial scale determine herbivore co-occurrence in a subtropical forest. *Ecological Monographs*, *92*(1). https://doi.org/10.1002/ecm.1492

Warren, D. L., Glor, R. E., & Turelli, M. (2008). ENVIRONMENTAL NICHE EQUIVALENCY VERSUS CONSERVATISM: QUANTITATIVE APPROACHES TO NICHE EVOLUTION. *Evolution*, *62*(11), 2868–2883. https://doi.org/10.1111/j.1558-5646.2008.00482.x

Webb, C. O. (2000). Exploring the Phylogenetic Structure of Ecological Communities: An Example for Rain Forest Trees. *The American Naturalist*, *156*(2), 145–155. https://doi.org/10.1086/303378

Webb, C. O., Ackerly, D. D., McPeek, M. A., & Donoghue, M. J. (2002). Phylogenies and Community Ecology. *Annual Review of Ecology and Systematics*, *33*(1), 475–505. https://doi.org/10.1146/annurev.ecolsys.33.010802.150448

Weiher, E., Freund, D., Bunton, T., Stefanski, A., Lee, T., & Bentivenga, S. (2011). Advances, challenges and a developing synthesis of ecological community assembly theory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1576), 2403–2413. https://doi.org/10.1098/rstb.2011.0056

Weiher, E., & Keddy, P. A. (1995). Assembly Rules, Null Models, and Trait Dispersion: New Questions from Old Patterns. *Oikos*, *74*(1), 159. https://doi.org/10.2307/3545686

Wiegmann, B. M., Trautwein, M. D., Winkler, I. S., Barr, N. B., Kim, J.-W., Lambkin, C., Bertone, M. A., Cassel, B. K., Bayless, K. M., Heimberg, A. M., Wheeler, B. M., Peterson, K. J., Pape, T., Sinclair, B. J., Skevington, J. H., Blagoderov, V., Caravas, J., Kutty, S. N., Schmidt-Ott, U., … Yeates, D. K. (2011). Episodic radiations in the fly tree of life. *Proceedings of the National Academy of Sciences*, *108*(14), 5690–5695. https://doi.org/10.1073/pnas.1012675108

Wilson, J. J. (2011). Assessing the Value of DNA Barcodes for Molecular Phylogenetics: Effect of Increased Taxon Sampling in Lepidoptera. *PLoS ONE*, *6*(9), e24769. https://doi.org/10.1371/journal.pone.0024769

Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., Dormann, C. F., Forchhammer, M. C., Grytnes, J., Guisan, A., Heikkinen, R. K., Høye, T. T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M., Normand, S., Öckinger, E., Schmidt, N. M., … Svenning, J. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, *88*(1), 15–30. https://doi.org/10.1111/j.1469-185X.2012.00235.x

Yeates, D. K., Meusemann, K., Trautwein, M., Wiegmann, B., & Zwick, A. (2016). Power, resolution and bias: Recent advances in insect phylogeny driven by the genomic revolution. *Current Opinion in Insect Science*, *13*, 16–23. https://doi.org/10.1016/j.cois.2015.10.007