MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2013) doi: 10.1111/1755-0998.12173

OPINION

Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead

SIMON JOLY,*†[1] T. JONATHAN DAVIES,‡[1] ANNIE ARCHAMBAULT,§ ANNE BRUNEAU,* ALISON DERRY,¶ STEVEN W. KEMBEL,¶ PEDRO PERES-NETO,¶ JANA VAMOSI** and TERRY A. WHEELER††
*Institut de recherche en biologie végétale, Département de sciences biologiques, Université de Montréal, 4101 Sherbrooke East, Montréal, Quebec, Canada H1X 2B2, †Montreal Botanical Garden, 4101 Sherbrooke East, Montréal, Quebec, Canada H1X 2B2, ‡Biology Department, McGill University, 1205 Dr Penfield, Montréal, Quebec, Canada H3A 1B1, §Québec Centre for Biodiversity Science, 1205 Dr Penfield, Montréal, Quebec, Canada H3A 1B1, ¶Département des sciences biologiques, Université du Québec à Montréal, 141 Avenue du Président-Kennedy, Montréal, Quebec, Canada H2X 1Y4, **Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4, ††Department of Natural Resources Sciences, McGill University, Macdonald Campus, Ste. Anne de Bellevue, Quebec, Canada H9X 3V9

Abstract

Ten years after DNA barcoding was initially suggested as a tool to identify species, millions of barcode sequences from more than 1100 species are available in public databases. While several studies have reviewed the methods and potential applications of DNA barcoding, most have focused on species identification and discovery, and relatively few have addressed applications of DNA barcoding data to ecology. These data, and the associated information on the evolutionary histories of taxa that they can provide, offer great opportunities for ecologists to investigate questions that were previously difficult or impossible to address. We present an overview of potential uses of DNA barcoding relevant in the age of ecoinformatics, including applications in community ecology, species invasion, macroevolution, trait evolution, food webs and trophic interactions, metacommunities, and spatial ecology. We also outline some of the challenges and potential advances in DNA barcoding that lie ahead.

Keywords: cryptic diversity, DNA barcoding, ecoinformatics, ecophylogenetics, food webs, intraspecific variation, macroevolution, phylogeny

Received 15 June 2013; revision received 3 September 2013; accepted 16 September 2013

Introduction

It has been 10 years since DNA barcoding was proffered as a rapid and effective means to identify species and to assess primary biodiversity (Hebert *et al.* 2003). Since then, large amounts of DNA barcode data have been accumulating in publicly available databases. It was once asked who would eventually use this wealth of information (Rubinoff *et al.* 2006), and a recent study suggested that DNA barcoding may be on the brink of irrelevancy (Taylor & Harris 2012). However, it seems obvious to us that DNA barcoding, as well as barcode data, has proven useful in many contexts (Valentini *et al.* 2009; Kress & Erickson 2012). Yet to date, applications have remained primarily focused on describing and documenting diversity, and until now, there has been little discussion on the potential variety of uses of barcoding data in ecology (but see Valentini *et al.* 2009 for a somewhat broader perspective). Here, we highlight new applications for barcode data relevant to ecology in the age of bioinformatics. For convenience, we group the potential applications of DNA barcode data into the following categories: (i) phylogenetic insights into the structure, assembly and function of communities, and macroevolutionary trends; (ii) cryptic diversity information relevant to ecological, abiotic and biotic interactions as well as to microbial communities; (iii) intraspecific variation and (iv) DNA barcoding metadata. Our classification is not meant to be exclusive, and as such, we depict the link between approaches, disciplines and applications as a

Correspondence: Simon Joly, Fax: +1 514 343 2288; E-mail: simon.joly@umontreal.ca

[1]Equal contributions; other authors in alphabetical order.

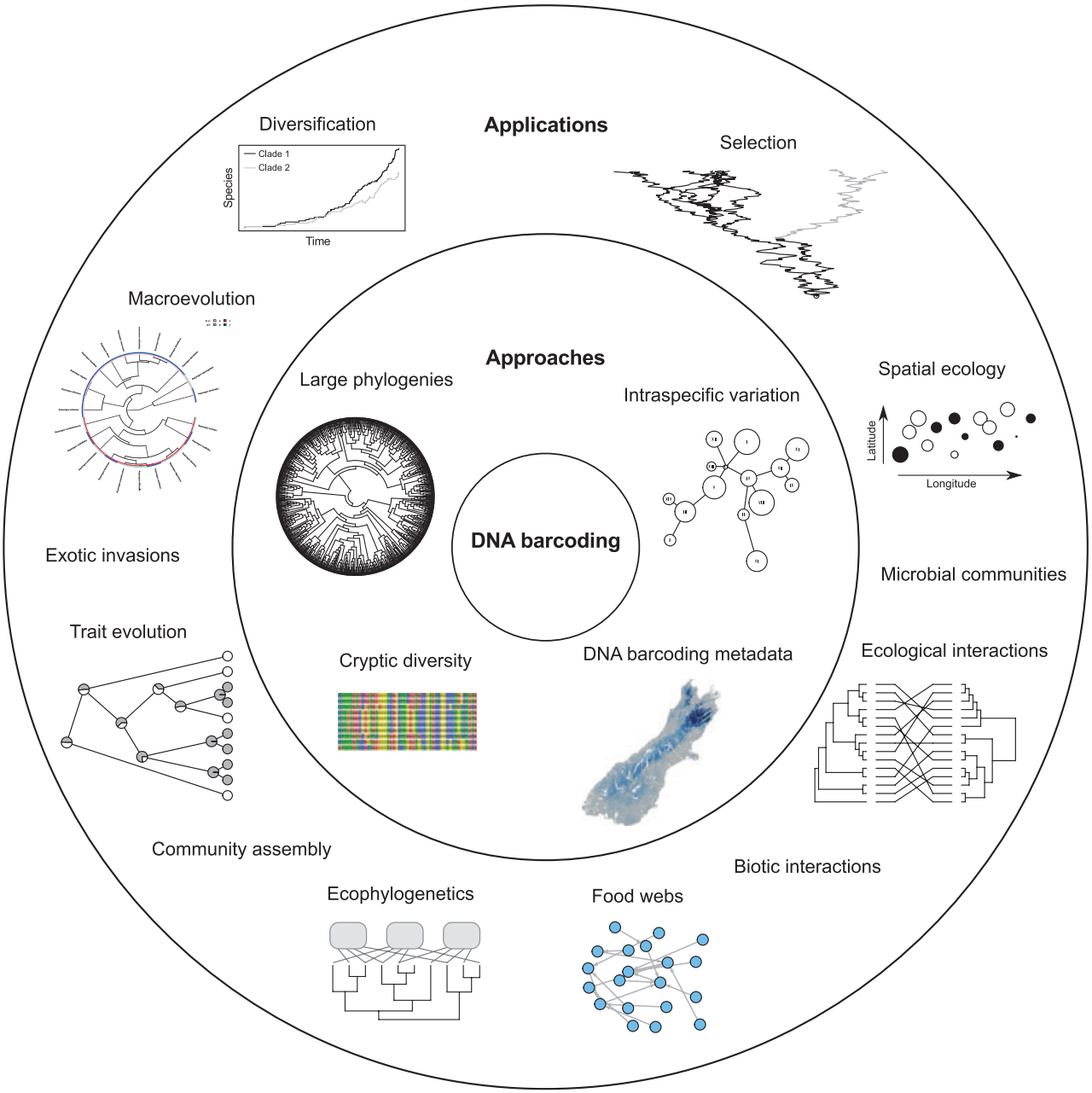

Fig. 1 Graphical representation of potential ecological applications of DNA barcoding, in the context of approaches and disciplines most relevant to DNA barcoding data.

series of concentric rings (Fig. 1), where the match between them will depend on the question of interest.

Background

DNA barcoding

DNA barcoding uses a short DNA fragment (the barcode marker) to distinguish species (Hebert *et al.* 2003). The approach requires a well-curated database that contains sequences from correctly identified individuals. The requirement for referenced specimens has sometimes created a philosophical division between DNA barcoding approaches in multicellular and unicellular organisms. Indeed, although sequence-based identification has been used for some time in bacteria and Archaea (Pace 1997), the fact that many unicellular organisms are known only from their sequences (i.e. there are no reference specimens), compounded by the difficulty in applying traditional species

concepts in these groups, likely explains why such studies are rarely acknowledged in the DNA barcoding literature.

Central to the success of DNA barcoding is the need for appropriate markers that would show sufficient variation yet have the potential to be amplified universally with general PCR primers. While marker choice rapidly converged on *CO1* for animals (Hebert *et al.* 2003) and on 16S rRNA genes for prokaryotes (Pace 1997), the decision to use *rbcLa* and *matK* in plants followed a long and sometimes heated debate (CBOL Plant Working Group *et al.* 2009). In fungi, the choice of barcode gene is still being discussed (Liu *et al.* 2012; Schoch *et al.* 2012).

As of October 2013, the Barcode of Life Data System (BOLD: www.boldsystems.org; Ratnasingham & Hebert 2007) had a total of 192 350 species and 2 509 708 eukaryote specimens with DNA barcode data. The vast majority of barcodes in BOLD are for animals (193 664 species with *CO1* sequences), followed by plants (54 974 species with *rbc*L sequences) and fungi (4266 species with ITS sequences), while the diverse unicellular Eukaryotes are still underrepresented (4999 species with CO1 sequences and 14 with ITS sequences; Fig. 2). DNA barcodes for prokaryotes are curated in a number of separate databases that contain several million sequences (e.g. Cole *et al.* 2009; Quast *et al.* 2012).

New uses for large phylogenies

The sheer size of the species-level phylogenies made possible using easily alignable barcode sequences offers opportunities to address ecological questions that until recently appeared intractable. In addition to the obvious potential that barcode phylogenies offer to the systematics community, large barcode phylogenies also allow us to explore, among many other questions, the assembly of ecological communities, macroevolutionary processes such as the tempo of trait evolution, extinction risk, and the contribution of phylogenetic history to ecosystem integrity. We outline below how large barcode phylogenies can be applied to these questions, but first, we will briefly describe what DNA barcoding can offer to the reconstruction of large phylogenies.

The key contribution of DNA barcode data to phylogeny reconstruction is in the completeness of taxonomic sampling. For many types of studies, it is important to be able to reconstruct phylogenetic relationships of all species from large groups of organisms (families, orders, etc.), because good species representation is often imperative for accurate and robust results when applying phylogenetic analyses to address ecological questions (e.g. Harmon *et al.* 2003; Kress *et al.* 2009; Cusimano *et al.* 2012). Good taxon sampling is even more important in comparative analyses, especially when these rely on parameter estimates from evolutionary models. For instance, Boettiger *et al.* (2012) has shown that large sample sizes are critical to detect balancing selection, and Nee *et al.* (1994) noted that incomplete taxon sampling can give a false impression of reduction in rates of diversification over time.

Of course, because a given gene tree might not always reflect the evolutionary history of species (Maddison 1997; Degnan & Rosenberg 2006), relying on a phylogeny reconstructed using a single or a few short markers is not recommended (particularly if they all are from the same organelle genome). In addition, the short length of barcode markers limits the phylogenetic information they contain, such that they might fail to resolve relationships among rapidly diverging lineages or between clades deeper in the phylogenetic tree, where there have likely been multiple substitutions per site. Therefore, even though empirical DNA barcode derived phylogenies can sometimes provide good hypotheses of taxon relationships (see Kress *et al.* 2009), the best approach is to combine DNA barcode data with other markers, preferably from independently evolving regions of the nuclear genome. When available, several markers should thus be analysed for as many taxa as possible to provide good support and accurate relationships among taxa, but the addition of barcode data could allow the inclusion of taxa for which this is the only information available and thus provide increased taxon representation. This approach is supported by studies based on simulated or empirical data that have shown that adding taxa with at least some missing sequence information (such as when only DNA barcode regions are available for some taxa) can significantly improve phylogenetic reconstruction, even where the proportion of missing data is high (Wiens 2005; Wiens & Tiu 2012). Moreover, increased sample sizes have been shown to result in more accurate phylogenetic estimates (Pollock *et al.* 2002), which is important for many downstream analyses that require an accurate topology and precise branch length estimates.

Phylogenetic reconstructions represent hypotheses of evolutionary relationships. As such, no single tree should be assumed to be true, irrespective of the amount of data employed in its reconstruction. Critically, when phylogenies are reconstructed from DNA data, such as DNA barcodes, it is possible to generate an estimate of the phylogenetic uncertainty (e.g. trees sampled from a Bayesian MCMC search or from a bootstrapping procedure), which can in turn be incorporated into the downstream analyses to assess how sensitive the conclusions are in relation to the uncertainties regarding the estimated tree. For most ecological applications

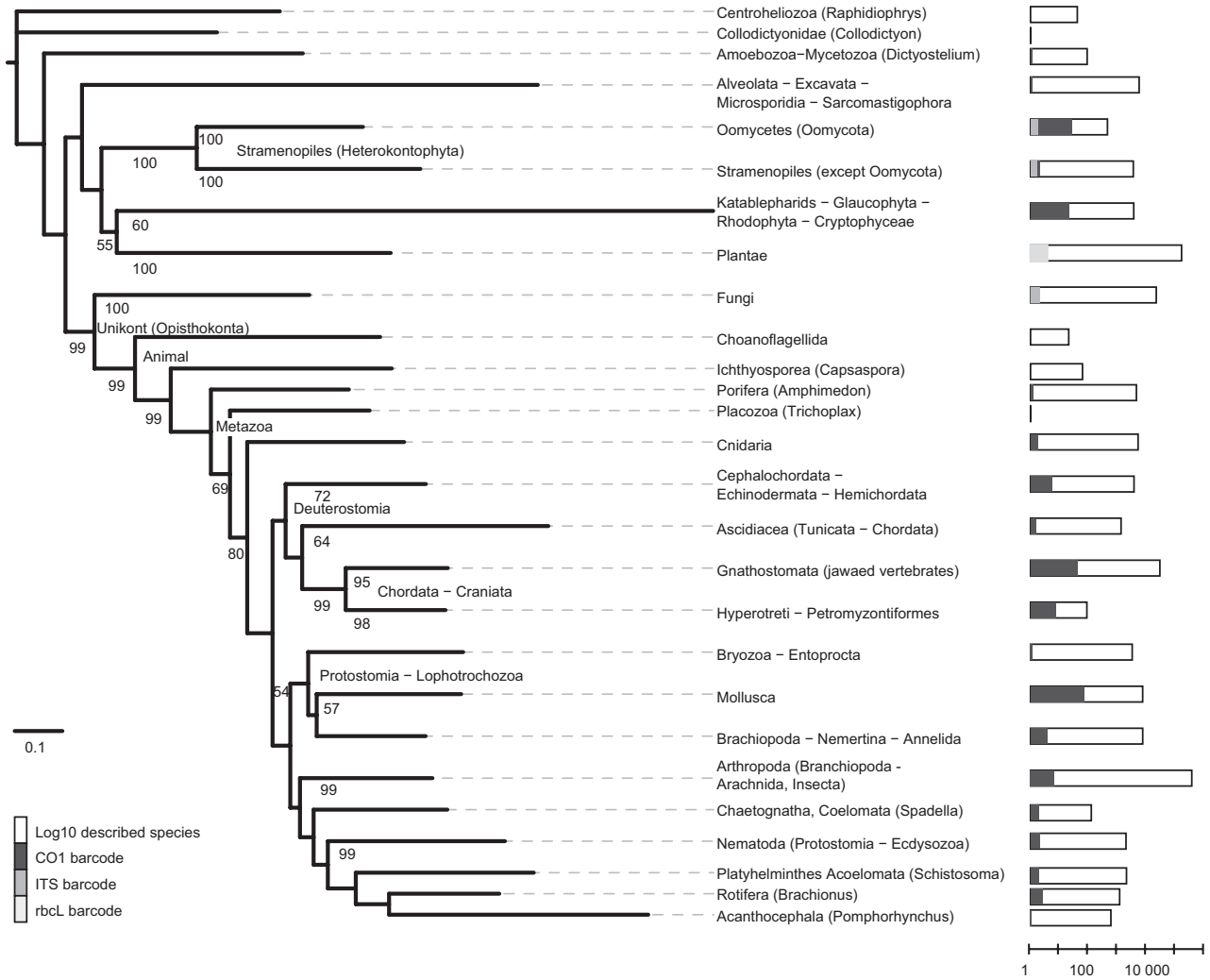

Fig. 2 For each major lineage of the eukaryote tree of life, the proportion of species with a known barcode (solid bar) is presented relative to the total number of described species (hollow bar). Species richness is $\log_{10}$ transformed. The tree was built by maximum likelihood from 29 amino acid gene sequences for 94 species (Appendix S1). Several species were collapsed to single, higher-level clades for ease of interpretation. The complete figure (Fig. S1) and data set (Table S1) are available as supporting information.

using phylogenetic trees, knowing the 'true' tree is not a prerequisite. What is important is to have comprehensive taxonomic sampling and to be able to include phylogenetic uncertainty into the statistical analyses. For example, Moore & Donoghue (2009) have taken into account the uncertainty in tree topology and divergence times to conclude that single-seeded fruits in *Viburnum* are associated with higher rates of diversification than other species of the family Adoxaceae. In contrast to trees reconstructed from DNA sequences, 'opinion-based' phylogenies provide no estimation of branch lengths and little information regarding phylogenetic uncertainty.

In summary, DNA barcode data provide an affordable, convenient and rapid way to include within phylogenetic analyses the large numbers of organisms for which no other sequence information is available.

Ecophylogenetics

Community assembly. By using universal primers, DNA barcode approaches enable the generation of complete phylogenetic hypotheses for entire communities of organisms (e.g. trees: Kress *et al.* 2009; Bacteria: Barbern & Casamayor 2010; Wang *et al.* 2012; Coleoptera: Baselga *et al.* 2013), allowing ecologists to address fundamental questions relating to the processes underlying their distribution and assembly (Webb *et al.* 2002; Mouquet *et al.* 2012). Most studies using such 'ecophylogenetic' approaches are predicated on the observation that there are often relationships between the ecological similarity of species and their phylogenetic relatedness (Cahill *et al.* 2008). Under this assumption, patterns of phylogenetic relatedness in local communities compared with the

patterns expected under null models of community assembly can provide insights into the relative importance of various processes such as dispersal, competition, filtering or drift during community assembly (Emerson & Gillespie 2008; Vamosi *et al.* 2009). While distinguishing among the signatures of different processes may be problematic (Mayfield & Levine 2010), for example, competition versus environmental filtering, several studies have demonstrated that the relative importance of niche-based versus species-neutral assembly processes can be assessed by incorporating phylogenetic information into analyses of community structure (Kembel 2009; Mouquet *et al.* 2012; Peres-Neto *et al.* 2012). The inclusion of DNA barcode data in such ecophylogenetic studies is now starting to become more common (e.g. Swenson 2012). In some instances, barcodes also allow the investigation of diversity patterns below the species level, such as at the level of genes and genealogies (see Baselga *et al.* 2013).

Ecological invasions. Invasive species are increasingly being recognized as a major threat to ecosystems (Pejchar & Mooney 2009). Although documented extinctions are rare, invasives have been linked to declines in native species (Mooney & Cleland 2001), and their economic impacts are potentially large (Pejchar & Mooney 2009). Predicting which species pose a threat of invasion following introduction is a crucial step for their management although it remains a major challenge. The importance of predicting invasiveness is particularly critical because logistics and funding for their control are limited even though early intervention minimizes total control costs. Two fundamental, but opposing processes have been linked to the success of invading species. First, introduced species may be more likely to become invasive in the absence of close relatives in the native species pool because competition with native species would be weaker – termed Darwin's naturalization hypothesis (Rejmánek 1996). Second, introduced species closely related to the native species pool may be more likely to become invasive because they share similar environmental preferences to the invaded community. Darwin recognized both possibilities, and this conflict has sometimes been referred to as Darwin's naturalization conundrum (Diez *et al.* 2008). A growing number of studies have explored the phylogenetic distance between invasive and native species although results have been mixed (Thuiller *et al.* 2010). Several possible explanations for this seeming lack of agreement among studies have been suggested. First, the process of invasion might be different among different taxa and regions. Second, phylogeny (or our current estimates of phylogeny) might not always capture the important ecological similarities and differences between species, for example, key functional traits related to resource use might demonstrate only weak phylogenetic conservatism. Third, different processes might operate at different spatial scales (Thuiller *et al.* 2010). Fourth, phylogenetic distance may not scale similarly and linearly with invasiveness potential across species in the phylogeny. Future studies aimed at predicting invasive potential using phylogeny must therefore consider the phylogenetic distribution of key traits as well as the spatial scale of analysis; nonetheless, the power of phylogenetic approaches will rest upon the availability of accurate, broadly sampled and well-resolved phylogenetic trees. Critically, many 'metaphylogeny' approaches, in which missing species are pasted into an existing phylogenetic framework, often reach their limit of resolution at the genus level and represent species relationships as polytomies, which is problematic given that invasives are often congeneric with native species. In such cases, large regional barcode-augmented phylogenies are likely to allow for more powerful tests by covering a greater breadth of taxa and helping resolve relationships among more closely related taxa.

Macroevolution

Macroevolution, broadly defined as evolutionary processes above the species level, captures the radiation (and contraction) of lineages and their associated morphological diversification. The fossil record sheds some light on macroevolutionary trends and the major evolutionary transitions in the history of life on Earth, including mass extinctions (Raup & Sepkoski 1982) and the rise to dominance of flowering plants (Crane & Lidgard 1989). However, because of taxonomic biases in preservation (i.e. taphonomic bias), fossil data will always present an incomplete and patchy record. Phylogenetic methods allow us to reconstruct the evolutionary histories of character traits and estimate ancestral states using information from extant taxa (e.g. Huelsenbeck *et al.* 2003; Pagel *et al.* 2004). Such information is integral to phylogenetic classification systems that define clades based on synapomorphies (shared derived traits), thereby forming the foundation of our understanding of biodiversity and comparative biology. Many methods of ancestral reconstruction stress the importance of nearly complete sampling (Barraclough & Nee 2001; but see FitzJohn *et al.* 2009). For this reason, most studies that have explored evolutionary transitions have until now been focused on traits that vary at lower taxonomic levels (e.g. family and below) using group-specific phylogenetic markers. However, as more barcode data become available, it will be possible to extend these studies to larger, more inclusive, taxonomic groups and to look for common evolutionary trends or investigate evolutionary processes that could have occurred over longer periods of time. Supermatrix methods have already demonstrated

the power of broad taxonomic sampling, for example, in linking rates of molecular evolution and plant life histories (Smith & Donoghue 2008). The inclusion of barcode data will allow us to simultaneously increase taxonomic depth of sampling.

Trait evolution. Initially, the introduction of phylogenetic methods in ecology aimed to control for phylogenetic nonindependence when testing for association between traits, or between traits and the environment (Felsenstein 1985). Many such phylogenetic comparative methods are now available (e.g. see Freckleton *et al.* 2002; Smith 2010). Phylogenies with good taxon representation can contribute to comparative biology by providing an opportunity to unravel the phylogenetic history of morphological evolution that could help address large-scale questions related to the ecological characteristics of species. For instance, Price *et al.* (2011) used a time-calibrated phylogeny to show that fishes living in coral reefs have evolved functional morphological diversity twice as fast as non-reef species. Along similar lines, it is possible to use barcode-type phylogenies to test hypotheses related to ecological adaptation (Evans *et al.* 2009), adaptive radiations (Alfaro *et al.* 2009) or phylogenetic niche conservatism (see Savage & Cavender-Bares 2012 for an example on willow communities). In addition, it is possible to explore evidence for ecological adaptation by contrasting observed differences in ecological traits between sister lineages with expectations derived from neutral evolutionary models (e.g. using disparity through time plots (Harmon *et al.* 2003)). These analyses require well-sampled, time-calibrated phylogenetic trees, which would be greatly facilitated by the production of barcode data.

Diversification. The tempo of evolution is an additional aspect of phylogenetic history that can be tested using time-calibrated barcode-augmented phylogenies. Patterns of species radiations can be evaluated from phylogenies assuming a molecular clock calibrated from the fossil record. These may be visualized with lineage through time plots, and diversification rates can be estimated. Such approaches might help us understand the importance of diversification in shaping global biodiversity gradients (Mittelbach *et al.* 2007). For instance, recent studies indicate that historical events have been of paramount importance in shaping present-day species richness patterns (reviewed in Wiens *et al.* 2011). Yet, phylogenetic estimations of speciation and extinction rates along latitudinal gradients have found inconsistent evidence for increased speciation rates in tropical regions (Weir & Schluter 2007; Jansson & Davies 2008). If similar speciation rates between tropical and temperate regions were confirmed, it would suggest that gradients in diversity are driven by longer times for speciation in tropical climates (Wiens & Donoghue 2004; Perret *et al.* 2007). In addition, the idea of diversity-dependent diversification is now maturing (Rabosky 2009; Vamosi & Vamosi 2010), with some indications that ecological limits might be higher in tropical climes (Rabosky 2009). Comparing phylogenetic diversity patterns (e.g. phylogenetic diversity and phylogenetic betadiversity) between temperate and tropical latitudes provides one novel way forward and helps connect local processes with large-scale biogeographical patterns (Kooyman *et al.* 2011; Davies & Buckley 2012). It is important to note that asymmetries in sampling and taxonomic expertise between regions limit the utility of current metaphylogenies, making barcodes particularly relevant for such large-scale biogeographical comparisons. Improved species-level phylogenetic trees in combination with new statistical methods that allow the simultaneous estimation of speciation and extinction rates that are associated with transitions in discrete (Maddison *et al.* 2007) and quantitative (FitzJohn 2010) character states and/or geographical locations (Goldberg *et al.* 2011) will greatly improve our understanding of the influence of diversification rates in shaping global biodiversity patterns.

Cryptic diversity

Barcoding approaches are especially useful for investigating cryptic diversity (e.g. resource diversity within trophic interactions, microbial communities, and intraspecific variation) and barcoding-type data have long been used to study microbial communities (Pace 1997). The advantages of barcoding to study microbial, environmental or other cryptic (e.g. plant endophytes, soil invertebrates, diverse arthropod taxa) communities have been widely recognized (reviewed in Valentini *et al.* 2009), but it is possible to go much beyond primary diversity assessments and further investigates ecological aspects associated with cryptic diversity. In the following paragraphs, we describe how the investigation of cryptic diversity via DNA barcodes can shed a new light on ecological interactions.

Food webs

Evolutionary studies of mutualism, parasitism and trophic cascades could benefit greatly from barcode data, where much of biodiversity is either cryptic, microscopic or both (Smith *et al.* 2011). Some of the benefits to barcode data involve the classic barcoding arguments (e.g. Valentini *et al.* 2009) based upon better estimates of diversity because the network of trophic interactions can then be obtained with greater speed and accuracy, as well as at larger scales. But often, DNA barcoding provides information that could not be obtained otherwise,

such as the possibility to identify plant species eaten by leaf beetles by extracting the total beetle DNA (Jurado-Rivera *et al.* 2009). Moreover, the accuracy and low cost of DNA barcoding open new research avenues for the study of trophic interactions by allowing comparative studies of trophic networks in space and time (Smith *et al.* 2011). This has the potential to help ecologists to get a better understanding of the functioning of ecosystems. In many systems, the dominant flow of energy may be mediated through network links that have historically been understudied, such as parasitism (Lafferty *et al.* 2006), and as such DNA barcoding could offer interesting solutions for describing the flow of energy through these pathways.

In addition, the universality of barcode methods facilitates the reconstruction of phylogenetic trees for multiple trophic levels within interaction webs (e.g. plant–pollinator interaction webs). By matching the branching patterns between trees, it is possible to disentangle patterns of covicariance (whereby hosts and parasites have experienced similar biogeographical histories) versus co-evolution (whereby hosts have adapted to closely related species) at a scale much larger than previously possible (Smith *et al.* 2008).

Ecological interactions

DNA barcoding approaches have the potential to provide new insights into species interactions. For instance, with DNA barcoding it is possible to investigate 'hidden' aspects of biodiversity by extracting DNA from environmental samples and potentially get more information than available from traditional (morphological) biodiversity assessments. Recent examples of such applications have included the characterization of belowground plant diversity by Kesanakurti *et al.* (2011) and Hiiesalu *et al.* (2012). These two studies found a correlation between the aboveground and the belowground plant species diversity, but both observed a greater diversity belowground than aboveground. Interestingly, the response to environmental factors such as soil fertility had inverse effects above and below ground: higher soil fertility decreased above ground plant species diversity while it increased the belowground plant species diversity (Hiiesalu *et al.* 2012). Clearly, such applications have the potential for allowing a better understanding of how biodiversity responds to environmental factors. Similar barcoding approaches also have potential to help understand the competitive dynamics between neighbouring plants at the level of root interactions, a field that has traditionally faced significant methodological challenges (Tosti & Thorup-Kristensen 2010). These represent specific examples, but the same type of questions could certainly be applied to other organisms or systems as well. Nonetheless, major challenges remain, for example differentiating between living and dead material (where this is important), and inferring interaction strengths from co-occurrence data obtained from DNA barcode identifications.

Microbial communities

Microbial ecologists have made extensive use of molecular sequence data to study the structure of microbial communities. The popularity of sequence data among microbiologists has probably been driven by the fact that direct observation of the organisms in these communities is challenging, but also because of the availability of barcode data and tools for studying microbial communities (Pace 1997). While these microbial studies are not commonly referred to as 'barcoding' studies, the approach is more or less equivalent, and the methods and resources developed for molecular analysis of microbial communities provide an example of how barcoding data sets could be applied to macroorganismal biology in the future. The most commonly used barcode gene in studies of bacterial and archaeal communities has been the small subunit ribosomal 16S rRNA. Countless studies have used this marker as a barcode to quantify microbial community structure based on DNA sequences from environmental samples (Hugenholtz *et al.* 1998). Such studies suggest that we have only begun to scratch the surface of microbial biodiversity; for example, of the millions of bacterial species estimated to exist (Curtis *et al.* 2002), only a tiny fraction have been described scientifically or captured in existing sequence databases (Wu *et al.* 2009).

Investigating cryptic diversity through DNA barcoding allows researchers to address ecological questions on several fronts. Molecular barcoding of microbial communities has revealed that microbes exhibit similar patterns of geographical structure as that reported for macroorganisms (Martiny *et al.* 2006; Nemergut *et al.* 2011). Other studies have investigated the role of microorganisms in ecosystem function (Torsvik & Øvreås 2002; Van Der Heijden *et al.* 2008). Specific examples involve the description of microorganisms communities that are associated with host health and fitness in humans (Turnbaugh *et al.* 2008; Qin *et al.* 2010; The Human Microbiome Project Consortium 2012), as well as with bioremediation capacity of diesel contaminated soils in the arctic (Yergeau *et al.* 2012).

Intraspecific variation

While the focus of many published barcode studies has been on the detection of interspecific variation, intraspecific variation in DNA barcodes has untapped potential for ecological applications. The main

advantages reside in the relative ease with which studies can be compared across multiple species at the community and metacommunity levels. In contrast, information-rich population genetic markers such as microsatellites are species specific and both labour-intensive and costly to develop for nonmodel species (Selkoe & Toonen 2006). Here, we summarize previously overlooked ecological applications of intraspecific variation that can be also captured by barcoding tools, but additionally highlight some of their limitations.

Spatial ecology

Metacommunities are defined as groups of interacting species in spatially separated habitat patches connected by dispersal of individuals (Hanski & Gaggiotti 2004). Considering that barcode markers are generally thought to evolve neutrally (but see below), their variation should be informative of the spatial and historical features of landscapes through drift and migration. Our knowledge of ecological and evolutionary processes within metacommunities has been particularly limited by technical challenges in measuring dispersal (see Jacobson & Peres-Neto 2010 for a review), and barcoding is opening new opportunities in this direction. For example, Craft *et al.* (2010) used population genetics of DNA barcodes from numerous species of tropical Lepidoptera in the same community to assess the generality of host-associated population genetic structure and the degree of isolation by geographical distance. Importantly, they found that this broad sampling approach was essential for detecting overall trends at the level of the ecological community. Baselga *et al.* (2013) also used DNA barcoding to reveal broad spatial patterns of intraspecific genetic diversity in whole communities to compare levels of biodiversity across levels of biological organization in aquatic beetles (genetic and community diversity). Spatial models (Leibold *et al.* 2010; Thomassen *et al.* 2010; Peres-Neto *et al.* 2012) can subsequently be applied to intraspecific barcoding of communities to evaluate how landscape processes (selection, biogeography and habitat connectedness and proximity) affect coexisting species across landscapes while bypassing assumptions of neutrality and equilibrium, which are often violated, but typically assumed in population genetic models. Population-level comparisons at the community level can therefore provide a first step towards understanding processes underlying diversification and specialization in an ecological context and reveal patterns not detectable by single-species studies.

Selection

It is important to appreciate that DNA barcoding markers that are located in the genomes of organelles (as for plants and animals, but not fungi) are generally transmitted uniparentally as a single unit (i.e. without recombination among genes). Therefore, although organelles are expected to evolve quasi-neutrally (Manel *et al.* 2003; Storfer *et al.* 2007, 2010), selection at any position in the organelle's genome will affect variation at the barcode marker. Because (undetected) selection can overwhelm neutral genetic signals, the use of organelle DNA as a population genetic tool has met substantial criticism (Galtier *et al.* 2009). For this reason, it is important to proceed with care when interpreting genetic patterns obtained from barcode data. At the same time, if a gene in the mitochondria is under selection, barcode markers might then provide a powerful tool for investigating the genetic consequences of environmental change (Derry *et al.* 2009; Sork *et al.* 2010). Further, because DNA barcodes can be variable among populations, it could be feasible to find polymorphisms correlated with a given trait, providing insights into species eco-evolutionary dynamics (Johnson & Stinchcombe 2007).

Limitations

The utility of DNA barcode data for understanding patterns of intraspecific variation, however, can be limited by the level of rigour in which data sets are collected, by low mutation rates of barcodes in certain organisms and by differences between gene trees and species trees. First, the design of many barcoding studies in terms of within-species sampling is insufficient for inferring information about within-species variation as often <10 individuals are used per species to construct DNA barcoding libraries. Therefore, to benefit from the use of barcodes for inferring how population genetic structure in metacommunities relates to ecological phenomena, we advocate that studies must be explicitly designed to adequately capture within-species genetic variation and ecological information across landscapes (e.g. Craft *et al.* 2010). A second limitation of barcodes for comparative population genetic studies in metacommunities is slow mutation rates at the intraspecific level for some organisms (e.g. plants). Given these slower mutation rates, barcodes at the intraspecific level are most often used to understand historical processes that occur in populations across broad spatial scales (Manel *et al.* 2003; Storfer *et al.* 2007, 2010), and molecular markers such as microsatellites are more useful for providing resolution of recent and on-going processes related to gene flow and drift across fine spatial scales. A logical follow-up step for studies that use barcodes to explore broad-scale ecological patterns in intraspecific variation across multispecies assemblages would be to incorporate additional data from nuclear loci and to test explicit hypotheses relating to the causes and consequences of genetic divergence.

A third limitation of barcode data at the intraspecific level is that a DNA barcode represents a single evolutionary marker. The evolutionary history contained in the barcode marker therefore represents a single outcome of the stochastic process of genetic drift in finite populations. Consequently, genealogies derived from barcode data might not always reflect the evolutionary history of organisms, a problem particularly acute in rapidly diverging groups or populations (Degnan & Rosenberg 2006), but less problematic when longer, macroevolutionary, timescales are considered. Nevertheless, despite these limitations, we suggest that tremendous information could be gleaned about the distribution and structure of genetic variation in the many different species in metacommunities so long as studies are carefully designed with sufficient sampling detail at the intraspecific level.

DNA barcoding metadata

An interesting aspect of the DNA barcoding effort is that it offers more than just sequence data. For instance, samples can be associated with metadata such as taxonomic information, geographical coordinates, elevation, reproductive strategies, etc., which can be deposited in databases such as the Barcode of Life Data (BOLD) System. These metadata seem to have been underutilized until now. One study has used DNA barcoding data and associated metadata from the birds of North America as a test case to assess conservation priorities using phylogenetic diversity measures (Guralnick & Hill 2009). In another study on North American fishes, April *et al.* (2013) barcoded 5674 individuals from nearly all (752) native freshwater North American fishes and used information on latitude, body size, metabolic rate and temperature to conclude that latitude and metabolic rates are the best predictors for intra- and interspecific genetic divergence. However, the potential uses for barcode metadata are much greater, and one could easily imagine using geographical coordinate to perform phylogeographical analyses that integrate geographical information. Geographical data for well-sampled taxa might also be useful for generating species distribution models, and predicting impacts of disturbance or climate change on species distributions, complementing data available in other electronic repositories, such as the Global Biodiversity Information Facility (GBIF). Photographic images, submitted with barcoded vouchers, contain an additional wealth of information. For example, in plants, image data might provide information on phenology, such as flowering times, allowing the documentation of phenological shifts over space and time, as has been explored with herbarium data (e.g. Miller-Rushing *et al.* 2006). Because metadata related to DNA barcoding projects are accumulating rapidly and that these are often stored in different databases (e.g. Hemery *et al.* 2013), one ongoing objective in the scientific community is to link together databases that have complementary information to maximize the use of this wealth of information (Parr *et al.* 2012).

Perspectives

The enormous amount of data that has been accumulating since the launch of DNA barcoding (Hebert *et al.* 2003; Fig. 2) represents incredible opportunities for ecologists to address new and inspiring questions that were beyond the remit of traditional data collection efforts. However, DNA barcoding is likely to face important challenges in the future.

First, while the volume of DNA barcode data has been accumulating at an increasing rate, it is nonetheless important to recognize that many taxa remain underrepresented. Such taxa should become high priority for future collection efforts. In addition, multiple records per species are required to explore intraspecific variation and landscape genetics. Despite advances in sequencing technology, the challenges of collecting such data are vast, and beyond the current research capacity. The generation of a comprehensive DNA barcode library will require the coordinated contributions of multiple researchers, specializing in different taxa or regions.

One of the factors in the success of the International Barcode of Life (www.ibol.org) was a funding structure that contributed to subsidizing sequencing costs across an array of major projects; this has incentivized researchers to work towards a common goal of building a comprehensive barcode database. Although the funding climate is unpredictable globally, the community should continue to strive to mobilize financial support for DNA barcoding initiatives. The erosion of low-cost DNA barcoding would be a double blow, because researchers in the most biodiverse parts of the world might be those least able to afford full costs, and the value of existing barcode data increases as the number of barcoded samples increases.

Second, as data accumulate, data management and curation become increasingly important. For barcoding to be useful, databases need to be well curated and accessible. In addition, barcoding data need to be associated with information-rich metadata, preferably in association with equally well-curated specimen databases and collections. In science, money more often goes to data discovery than towards maintaining and managing already acquired data. Yet, this is a critical issue, and the added benefits of having well-curated databases are self-evident when we look at the existing databases for microbes (microbialgenomics.energy.gov/databases.shtml) and gene sequences (www.ncbi.nlm.nih.gov/genbank/). This challenge is not specific to barcoding although constant

accumulation of genomic data through next-generation sequencing applications will face the same problem.

Last, next-generation sequencing techniques represent both challenges and opportunities for barcoding in the future (Baird & Hajibabaei 2012; Taylor & Harris 2012; Poisot *et al.* 2013): opportunities because the increased output of these sequencing techniques allows ever more rapid and less expensive analysis of environmental samples; challenges because barcoding will have to adapt to the increasing ease of sequencing genomes or transcriptomes and to the inherent advantages of investigating genome wide variation. For example, in addition to targeted sequencing of microbial barcodes, it is now possible to sequence the metagenome – the genomes of all organisms present in a sample (Handelsman 2004). Yet, barcoding appears to be embracing these challenges, and next-generation 'meta-barcoding' studies are quickly accumulating (see Shokralla *et al*. 2012). Perhaps, until we can consider sequencing the complete genome of all living species, it may be that for ecologists and others, barcoding databases will remain useful for some time into the future.

Here, we have attempted to illustrate the tremendous potential of DNA barcoding to ecologists. Until now, DNA barcoding has been used mainly as a tool for biodiversity discovery. We have described above several additional applications of DNA barcode data, but there are certainly many other possible uses, for example, using barcode diversity as an indicator of ecosystem 'health' given growing evidence for a strong link between phylogenetic diversity and ecosystem functions (Cadotte 2013). We suggest that many ecologists might benefit by collecting vouchers in a manner consistent with successful amplification of barcode regions. To date, there has tended to be a separation between those collecting barcode data and ecologists who might benefit from using such data. We believe DNA barcoding research bears great promise, but to maximize the value of DNA barcode data, there must be increased collaboration between molecular biologists, ecologists, systematists and ecoinformaticians.

Acknowledgements

The authors would like to acknowledge the financial support of the Quebec Centre for Biodiversity Sciences (QCBS), the Institut de recherche en biologie végétale (IRBV) and the Biology Department of McGill University for supporting a workshop that led to this review.

References

Alfaro ME, Santini F, Brock C *et al.* (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences, USA*, **106**, 13410–13414.

April J, Hanner RH, Mayden RL, Bernatchez L (2013) Metabolic rate and climatic fluctuations shape continental wide pattern of genetic divergence and biodiversity in fishes. *PLoS ONE*, **8**, e70296.

Baird DJ, Hajibabaei M (2012) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, **21**, 2039–2044.

Barbern A, Casamayor EO (2010) Global phylogenetic community structure and β-diversity patterns in surface bacterioplankton metacommunities. *Aquatic Microbial Ecology*, **59**, 1–10.

Barraclough TG, Nee S (2001) Phylogenetics and speciation. *Trends in Ecology & Evolution*, **16**, 391–399.

Baselga A, Fujisawa T, Crampton-Platt A *et al.* (2013) Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nature Communications*, **4**, 1892.

Boettiger C, Coop G, Ralph P (2012) Is your phylogeny informative? Measuring the power of comparative methods. *Evolution*, **66**, 2240–2251.

Cadotte MW (2013) Experimental evidence that evolutionarily diverse assemblages result in higher productivity. *Proceedings of the National Academy of Sciences, USA*, **110**, 8996–9000.

Cahill JF Jr, Kembel SW, Lamb EG, Keddy PA (2008) Does phylogenetic relatedness influence the strength of competition among vascular plants? *Perspectives in Plant Ecology, Evolution and Systematics*, **10**, 41–50.

CBOL Plant Working Group, Hollingsworth PM, Forrest LL *et al.* (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences, USA*, **106**, 12794–12797.

Cole JR, Wang Q, Cardenas E *et al.* (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, **37**, D141–D145.

Craft KJ, Pauls SU, Darrow K *et al.* (2010) Population genetics of ecological communities with DNA barcodes: an example from New Guinea Lepidoptera. *Proceedings of the National Academy of Sciences, USA*, **107**, 5041–5046.

Crane PR, Lidgard S (1989) Angiosperm diversification and paleolatitudinal gradients in cretaceous floristic diversity. *Science (New York, N.Y.)*, **246**, 675–678.

Curtis TP, Sloan WT, Scannell JW (2002) Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences, USA*, **99**, 10494–10499.

Cusimano N, Stadler T, Renner SS (2012) A new method for handling missing species in diversification analysis applicable to randomly or nonrandomly sampled phylogenies. *Systematic Biology*, **61**, 785–792.

Davies TJ, Buckley LB (2012) Exploring the phylogenetic history of mammal species richness. *Global Ecology and Biogeography*, **21**, 1096–1105.

Degnan JH, Rosenberg NA (2006) Discordance of species trees with their most likely gene trees. *PLoS Genetics*, **2**, e68.

Derry AM, Arnott SE, Shead JA, Hebert PDN, Boag PT (2009) Ecological linkages between community and genetic diversity in zooplankton among boreal shield lakes. *Ecology*, **90**, 2275–2286.

Diez JM, Sullivan JJ, Hulme PE, Edwards G, Duncan RP (2008) Darwin's naturalization conundrum: dissecting taxonomic patterns of species invasions. *Ecology Letters*, **11**, 674–681.

Emerson BC, Gillespie RG (2008) Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology & Evolution*, **23**, 619–630.

Evans MEK, Smith SA, Flynn RS, Donoghue MJ (2009) Climate, niche evolution, and diversification of the "bird-cage" evening primroses (Oenothera, sections Anogra and Kleinia). *The American Naturalist*, **173**, 225–240.

Felsenstein J (1985) Phylogenies and the comparative method. *The American Naturalist*, **125**, 1–15.

FitzJohn RG (2010) Quantitative traits and diversification. *Systematic Biology*, **59**, 619–633.

FitzJohn RG, Maddison WP, Otto SP (2009) Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Systematic Biology*, **58**, 595–611.

Freckleton RP, Harvey PH, Pagel M (2002) Phylogenetic analysis and comparative data: a test and review of evidence. *The American Naturalist*, **160**, 712–726.

Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541–4550.

Goldberg EE, Lancaster LT, Ree RH (2011) Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology*, **60**, 451–465.

Guralnick R, Hill A (2009) Biodiversity informatics: automated approaches for documenting global biodiversity patterns and processes. *Bioinformatics*, **25**, 421–428.

Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews*, **68**, 669–685.

Hanski IA, Gaggiotti OE (eds) (2004) *Ecology, Genetics and Evolution of Metapopulations*, 1st edn. Elsevier, Burlington, Massachusetts.

Harmon LJ, Schulte JA, Larson A, Losos JB (2003) Tempo and mode of evolutionary radiation in Iguanian lizards. *Science*, **301**, 961–964.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, series B*, **270**, 313–321.

Hemery LG, Ameziane N, Eleaume M (2013) Circumpolar dataset of sequenced specimens of *Promachocrinus kerguelensis* (Echinodermata, Crinoidea). *ZooKeys*, **315**, 55–64.

Hiiesalu I, Öpik M, Metsis M *et al.* (2012) Plant species richness belowground: higher richness and new patterns revealed by next-generation sequencing. *Molecular Ecology*, **21**, 2004–2016.

Huelsenbeck JP, Nielsen R, Bollback JP (2003) Stochastic mapping of morphological characters. *Systematic Biology*, **52**, 131–158.

Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology*, **180**, 4765–4774.

Jacobson B, Peres-Neto PR (2010) Quantifying and disentangling dispersal in metacommunities: how close have we come? How far is there to go? *Landscape Ecology*, **25**, 495–507.

Jansson R, Davies TJ (2008) Global variation in diversification rates of flowering plants: energy vs. climate change. *Ecology Letters*, **11**, 173–183.

Johnson MTJ, Stinchcombe JR (2007) An emerging synthesis between community ecology and evolutionary biology. *Trends in Ecology & Evolution*, **22**, 250–257.

Jurado-Rivera JA, Vogler AP, Reid CAM, Petitpierre E, Gómez-Zurita J (2009) DNA barcoding insect–host plant associations. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 639–648.

Kembel SW (2009) Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. *Ecology Letters*, **12**, 949–960.

Kesanakurti PR, Fazekas AJ, Burgess KS *et al.* (2011) Spatial patterns of plant diversity below-ground as revealed by DNA barcoding. *Molecular Ecology*, **20**, 1289–1302.

Kooyman R, Rossetto M, Cornwell W, Westoby M (2011) Phylogenetic tests of community assembly across regional to continental scales in tropical and subtropical rain forests. *Global Ecology and Biogeography*, **20**, 707–716.

Kress WJ, Erickson DL (2012) *DNA Barcodes*. Humana Press, New York, New York.

Kress WJ, Erickson DL, Jones FA *et al.* (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences, USA*, **106**, 18621–18626.

Lafferty KD, Dobson AP, Kuris AM (2006) Parasites dominate food web links. *Proceedings of the National Academy of Sciences, USA*, **103**, 11211–11216.

Leibold MA, Economo EP, Peres-Neto P (2010) Metacommunity phylogenetics: separating the roles of environmental filters and historical biogeography. *Ecology Letters*, **13**, 1290–1299.

Liu K-L, Porras-Alfaro A, Kuske CR, Eichorst SA, Xie G (2012) Accurate, rapid taxonomic classification of fungal large-subunit rRNA genes. *Applied and Environmental Microbiology*, **78**, 1523–1533.

Maddison WP (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.

Maddison WP, Midford PE, Otto SP (2007) Estimating a binary character's effect on speciation and extinction. *Systematic Biology*, **56**, 701–710.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, **18**, 189–197.

Martiny JBH, Bohannan BJM, Brown JH *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, **4**, 102–112.

Mayfield MM, Levine JM (2010) Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology letters*, **13**, 1085–1093.

Miller-Rushing AJ, Primack RB, Primack D, Mukunda S (2006) Photographs and herbarium specimens as tools to document phenological changes in response to global warming. *American Journal of Botany*, **93**, 1667–1674.

Mittelbach GG, Schemske DW, Cornell HV *et al.* (2007) Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecology Letters*, **10**, 315–331.

Mooney HA, Cleland EE (2001) The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences, USA*, **98**, 5446–5451.

Moore BR, Donoghue MJ (2009) A Bayesian approach for evaluating the impact of historical events on rates of diversification. *Proceedings of the National Academy of Sciences, USA*, **106**, 4307–4312.

Mouquet N, Devictor V, Meynard CN *et al.* (2012) Ecophylogenetics: advances and perspectives. *Biological reviews*, **87**, 769–785.

Nee S, Holmes EC, May RM, Harvey PH (1994) Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions: Biological Sciences*, **344**, 77–82.

Nemergut DR, Costello EK, Hamady M *et al.* (2011) Global patterns in the biogeography of bacterial taxa. *Environmental microbiology*, **13**, 135–144.

Pace NR (1997) A molecular view of microbial diversity and the biosphere. *Science*, **276**, 734–740.

Pagel M, Meade A, Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology*, **53**, 673–684.

Parr CS, Guralnick R, Cellinese N, Page RDM (2012) Evolutionary informatics: unifying knowledge about the diversity of life. *Trends in Ecology & Evolution*, **27**, 94–103.

Pejchar L, Mooney HA (2009) Invasive species, ecosystem services and human well-being. *Trends in ecology & evolution*, **24**, 497–504.

Peres-Neto PR, Leibold MA, Dray S (2012) Assessing the effects of spatial contingency and environmental filtering on metacommunity phylogenetics. *Ecology*, **93**, S14–S30.

Perret M, Chautems A, Spichiger R, Barraclough TG, Savolainen V (2007) The geographical pattern of speciation and floral diversification in the neotropics: the tribe Sinningieae (Gesneriaceae) as a case study. *Evolution*, **61**, 1641–1660.

Poisot T, Péquin B, Gravel D (2013) High-throughput sequencing: a roadmap toward community ecology. *Ecology and Evolution*, **3**, 1125–1139.

Pollock DD, Zwickl DJ, McGuire JA, Hillis DM (2002) Increased taxon sampling is advantageous for phylogenetic inference. *Systematic biology*, **51**, 664–671.

Price SA, Holzman R, Near TJ, Wainwright PC (2011) Coral reefs promote the evolution of morphological diversity and ecological novelty in labrid fishes. *Ecology Letters*, **14**, 462–469.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS *et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, **464**, 59–65.

Quast C, Pruesse E, Yilmaz P *et al.* (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, **41**, D590–D596.

Rabosky DL (2009) Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. *Ecology Letters*, **12**, 735–743.
Ratnasingham S, Hebert PDN (2007) BOLD: The barcode of life data system Available from http://www.barcodinglife.org. *Molecular Ecology Notes*, **7**, 355–364
Raup DM, Sepkoski JJ (1982) Mass extinctions in the marine fossil record. *Science*, **215**, 1501–1503.
Rejmánek M (1996) A theory of seed plant invasiveness: The first sketch. *Biological Conservation*, **78**, 171–181.
Rubinoff D, Cameron S, Will K (2006) A genomic perspective on the shortcomings of mitochondrial DNA for "Barcoding" identification. *Journal of Heredity*, **97**, 581–594.
Savage JA, Cavender-Bares J (2012) Habitat specialization and the role of trait lability in structuring diverse willow (genus Salix) communities. *Ecology*, **93**, S138–S150.
Schoch CL, Seifert KA, Huhndorf S *et al.* (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences, USA*, **109**, 6241–6246.
Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology letters*, **9**, 615–629.
Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, **21**, 1794–1805.
Smith SD (2010) Using phylogenetics to detect pollinator-mediated floral evolution. *New Phytologist*, **188**, 354–363.
Smith SA, Donoghue MJ (2008) Rates of molecular evolution are linked to life history in flowering plants. *Science*, **322**, 86–89.
Smith CI, Godsoe WKW, Tank S, Yoder JB, Pellmyr O (2008) Distinguishing coevolution from covicariance in an obligate pollination mutualism: asynchronous divergence in Joshua tree and its pollinators. *Evolution*, **62**, 2676–2687.
Smith MA, Eveleigh ES, McCann KS, Merilo MT, McCarthy PC, Van Rooyen KI (2011) Barcoding a quantified food web: crypsis, concepts, ecology and hypotheses. *PLoS ONE*, **6**, e14424.
Sork VL, Davis FW, Westfall R *et al.* (2010) Gene movement and genetic association with regional climate gradients in California valley oak (Quercus lobata Née) in the face of climate change. *Molecular Ecology*, **19**, 3806–3823.
Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the "landscape" in landscape genetics. *Heredity*, **98**, 128–142.
Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496–3514.
Swenson NG (2012) Phylogenetic analyses of ecological communities using DNA barcode data. In: *DNA Barcodes* (eds Kress WJ, Erickson DL), pp. 409–419. Humana Press, New York, New York.
Taylor HR, Harris WE (2012) An emergent science on the brink of irrelevance: a review of the past 8 years of DNA barcoding. *Molecular Ecology Resources*, **12**, 377–388.
The Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature*, **486**, 207–214.
Thomassen HA, Cheviron ZA, Freedman AH, Harrigan RJ, Wayne RK, Smith TB (2010) Spatial modelling and landscape-level approaches for visualizing intra-specific variation. *Molecular Ecology*, **19**, 3532–3548.
Thuiller W, Gallien L, Boulangeat I *et al.* (2010) Resolving Darwin's naturalization conundrum: a quest for evidence. *Diversity and Distributions*, **16**, 461–475.
Torsvik V, Øvreås L (2002) Microbial diversity and function in soil: from genes to ecosystems. *Current opinion in microbiology*, **5**, 240–245.
Tosti G, Thorup-Kristensen K (2010) Using coloured roots to study root interaction and competition in intercropped legumes and non-legumes. *Journal of Plant Ecology*, **3**, 191–199.
Turnbaugh PJ, Hamady M, Yatsunenko T *et al.* (2008) A core gut microbiome in obese and lean twins. *Nature*, **457**, 480–484.
Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. *Trends in Ecology & Evolution*, **24**, 110–117.
Vamosi JC, Vamosi SM (2010) Key innovations within a geographical context in flowering plants: towards resolving Darwin's abominable mystery. *Ecology Letters*, **13**, 1270–1279.
Vamosi SM, Heard SB, Vamosi JC, Webb CO (2009) Emerging patterns in the comparative analysis of phylogenetic community structure. *Molecular Ecology*, **18**, 572–592.
Van Der Heijden MGA, Bardgett RD, Van Straalen NM (2008) The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters*, **11**, 296–310.
Wang J, Soininen J, He J, Shen J (2012) Phylogenetic clustering increases with elevation for microbes. *Environmental Microbiology Reports*, **4**, 217–226.
Webb CO, Ackerly DD, McPeek MA, Donoghue MJ (2002) Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, **33**, 475–505.
Weir JT, Schluter D (2007) The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science*, **315**, 1574–1576.
Wiens JJ (2005) Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Systematic Biology*, **54**, 731–742.
Wiens JJ, Donoghue MJ (2004) Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution*, **19**, 639–644.
Wiens JJ, Tiu J (2012) Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. *PLoS ONE*, **7**, e42925.
Wiens JJ, Pyron RA, Moen DS (2011) Phylogenetic origins of local-scale diversity patterns and the causes of Amazonian megadiversity. *Ecology Letters*, **14**, 643–652.
Wu D, Hugenholtz P, Mavromatis K *et al.* (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature*, **462**, 1056–1060.
Yergeau E, Sanschagrin S, Beaumier D, Greer CW (2012) Metagenomic analysis of the bioremediation of diesel-contaminated canadian high arctic soils. *PLoS ONE*, **7**, e30058.

A.A. performed the phylogenetic analyses. All authors contributed sections of the manuscript. S.J. and T.J.D. wrote a first draft and all authors substantially revised the manuscript.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Complete eukaryote tree of life built by maximum likelihood from 29 amino acid gene sequences representing the number of described species (hollow bar) and the number of species with a known barcode (solid bar), with species richness $\log_{10}$ transformed.

Table S1 Taxonomic information for the organisms used in the phylogenetic analysis, and number of species and number of species with barcode data for each lineage represented by an organism in the analysis.

Appendix S1 Phylogenetic analysis of major eukaryote lineages.