

DNA barcodes for ecology, evolution, and conservation

W. John Kress¹, Carlos García-Robledo^{1,2,3}, Maria Uriarte⁴, and David L. Erickson¹

¹ Department of Botany, MRC-166, National Museum of Natural History, Smithsonian Institution, PO Box 37012, Washington, DC 20013-7012, USA

² Department of Entomology, MRC-187, National Museum of Natural History, Smithsonian Institution, PO Box 37012, Washington, DC 20013-7012, USA

³ Laboratory of Interactions and Global Change, Department of Multitrophic Interactions, Institute of Ecology (INECOL), Carretera Antigua a Coatepec Number 351, El Haya, Xalapa, Veracruz 91070, Mexico

⁴ Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY 10027, USA

The use of DNA barcodes, which are short gene sequences taken from a standardized portion of the genome and used to identify species, is entering a new phase of application as more and more investigations employ these genetic markers to address questions relating to the ecology and evolution of natural systems. The suite of DNA barcode markers now applied to specific taxonomic groups of organisms are proving invaluable for understanding species boundaries, community ecology, functional trait evolution, trophic interactions, and the conservation of biodiversity. The application of next-generation sequencing (NGS) technology will greatly expand the versatility of DNA barcodes across the Tree of Life, habitats, and geographies as new methodologies are explored and developed.

DNA barcodes: what, why, and how

In reference to the universal coded labels found on commercial products, the term ‘DNA barcode’ is now commonly applied by biologists to a standardized short sequence of DNA that can be recovered and characterized as a unique identification marker for all species on the planet [1] (see [Glossary](#)). For many users of DNA barcodes, identification of an unknown sample by correctly matching a specific genetic marker to a reference sequence library is the primary goal ([Box 1](#)). However, DNA barcodes can also be applied as tools for addressing fundamental questions in ecology, evolution, and conservation biology, such as: how are species assembled in communities? What multispecies interactions occur in previously poorly known environments (e.g., soils)? And, where are the most evolutionarily rich habitats to be targeted for protection? These uses of DNA barcodes, which have only recently been considered and offer some of the most exciting prospects for using this new taxonomic tool, are the primary focus of this paper.

A DNA barcode, in its simplest definition, is one or few relatively short gene sequences taken from a standardized

portion of the genome and used to identify species. The use of DNA sequences for biological identifications is not new [2], but the concept of a ‘DNA barcode’ for quick and reliable species-level identifications across all forms of life, including animals, plants, fungi, and microorganisms, was first proposed just over a decade ago [3]. A short DNA sequence of 600 base pairs (bp) in the mitochondrial gene encoding cytochrome c oxidase subunit 1 (CO1; [3]) has been accepted as a practical, standardized, species-level DNA barcode for many groups of animals ([Figure 1](#)). Given that CO1 does not work as a DNA barcode in plants [4] and fungi [5], a concerted search was required for more effective gene regions for these major groups of organism ([Figure 1](#)). The DNA barcode loci now most commonly used for plants [a combination of plastid *rbcL*, *matK*, and *trnH-psbA* with nuclear internal transcribed spacer (ITS)] and fungi (nuclear ITS) may never be as efficient and successful in identification as CO1 in animals. However, even for some groups of animals (e.g., some invertebrates [6], amphibians, and reptiles [7]), CO1 works poorly and other gene regions are being used as DNA barcodes ([Figure 1](#)).

The introduction of any new method of analysis in science often brings some controversy and concern, which has been the case with DNA barcoding, especially in the field of taxonomy (e.g., [8]). In some taxa, DNA barcode markers were not as effective as first proposed (e.g., [9]). Plants, which have inherently lower rates of nucleotide substitution in mtDNA compared with animals [10], were especially problematic during the early stages of developing DNA barcodes. In addition, the concern that DNA barcodes will give poor results or faulty identifications because of the complications of ancestral polymorphisms, hybridization, and/or introgression certainly applies to both plants and animals [10,11]. These complications can be particularly acute in some groups of plant in which hybridization is widespread and pseudogenes in the nuclear genome are common [12]. More concerted work is needed on taxa with extensive hybridization to verify whether DNA barcodes can successfully provide accurate identifications across all species.

However, it may also be true that poor species resolution with DNA markers is due in part to our imperfect and variable definition of species across the major lineages of

Corresponding author: Kress, W.J. (kressj@si.edu).

Keywords: DNA barcodes; ecology; next generation sequencing; phylogenetics; taxonomy.

0169-5347/

© 2014 Published by Elsevier Ltd. <http://dx.doi.org/10.1016/j.tree.2014.10.008>

Glossary

Alpha and beta species diversity: alpha diversity is calculated as the total species diversity in any single site or unit. Beta diversity quantifies site-to-site variability in community composition.

Community assembly: ecological communities are created through the arrival, reproduction, and local extinction of individual species. Community assembly processes drive colonization and extinction dynamics, and range from entirely stochastic to entirely deterministic (e.g., competition or environmental filtering).

Community phylogeny: an evolutionary framework for co-occurring species in a community that is based on historical relations among the member taxa.

Cryptic species: a complex of species in which members are reproductively isolated from each other yet morphologically indistinguishable.

DNA barcode: in the strict sense, a DNA barcode is one or more short gene sequences taken from a standardized portion of the genome and used to identify species. In a broad sense, a DNA barcode is any DNA sequence used for identification at any taxonomic level.

DNA barcode library: a set of DNA barcode sequences compiled from species of known taxonomic origin used to compare and identify DNA barcode sequences recovered from unknown samples.

DNA metabarcoding: the use of NGS to identify multiple species in a sample using DNA barcodes.

Ecological forensics: a specialized version of DNA metabarcoding in which trophic or network interactions among species are resolved by genotyping complex mixtures of individuals. These mixtures may take the form of an individual organism in which parasites, mutualists, diet items, and symbionts are recovered and sequenced, or complex mixtures of propagules, such as pollen grains and seeds.

Functional trait: a measurable property, phenotype, or characteristics of an organism that may influence its fitness (i.e., survival, growth, reproduction, or dispersal).

Habitat filtering: the process where only species that carry a specific trait are able to successfully colonize a habitat.

Mini-barcode: a portion of a standard DNA barcode that is targeted for use in ancient or degraded samples in which the complete DNA barcode sequence may be unable to be recovered.

Morphospecies: species recognized by taxonomists based on their morphology, but not yet classified or named as novel species or varieties.

Next-generation sequencing (NGS; or massive parallel sequencing): sequencing methods that differ from Sanger sequencing in the ability to sequence thousands to millions of small DNA fragments simultaneously. Several related but distinct technologies are referred to as NGS, including Illumina, Roche454, and Ion Torrent, among others whose unifying similarity is that they are not based on the Sanger Di-Deoxy method.

Niche differentiation: the process by which different resource use or different niches allows the co-occurrence of phenotypically different species.

PCR: a method to target and copy a segment of DNA from among many sequences in a genome or sample. The result is billions of copies of the target sequence, which may then be analyzed to infer the exact nucleotide sequence.

Phylogenetic diversity: the sum of genetic distances connecting all taxa in a dated or molecular-clock scaled phylogeny. It focuses on the overall evolutionary divergence among taxa rather than species number (Simpsons Diversity) alone in a community or habitat.

Tree of Life: a phylogenetic reconstruction of living organisms that demonstrates the shared evolutionary origins and divergences of the included lineages.

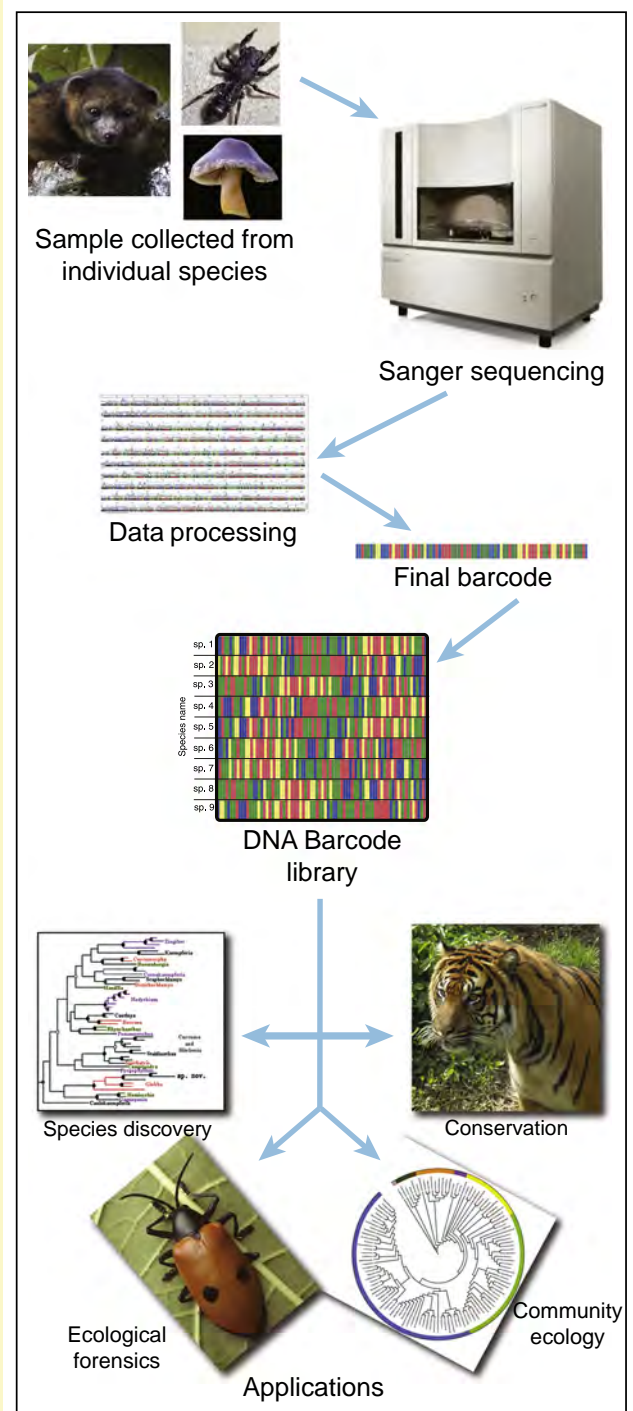
Trophic interactions: a network of interactions among autotrophs, predators, and their prey.

Box 1. Building the DNA barcode library using Sanger sequencing

The workflow for generating DNA barcodes for individual species entails two basic steps: (i) building the DNA barcode library of known species; and (ii) matching the DNA barcode sequence of an unknown sample against the barcode library for identification (Figure 1).

The DNA barcode library is a collection of DNA sequences associated with verified taxonomic identification and ideally with voucher specimens. A comprehensive DNA barcode library, which will be the most useful for multiple applications, is a recognized limiting factor because of the overwhelming number of species of plants, animals, and fungi already described by taxonomists. In this respect, museum specimens are a critical source of tissue for generating DNA barcodes with known vouchers. Nearest neighbor algorithms are usually used to assign an unknown sample to a known species by finding the closest database sequence to the

sample sequence [76]. Basic Local Alignment Search Tool (BLAST) is a common matching tool, provided through NCBI, that searches for correspondence between a query sequence and a sequence library [77].



TRENDS in Ecology & Evolution

Figure 1. Basic workflow for generating DNA barcodes using Sanger sequencing.

life. Even those who profess little confidence in DNA barcoding accept that our species concepts are not universal and often inconsistent among taxonomists [8]. In some taxa, the numbers of species recognized by different taxonomists may vary by up to 50% [8]. Although we have no comparative evidence across phyla that indicates that all

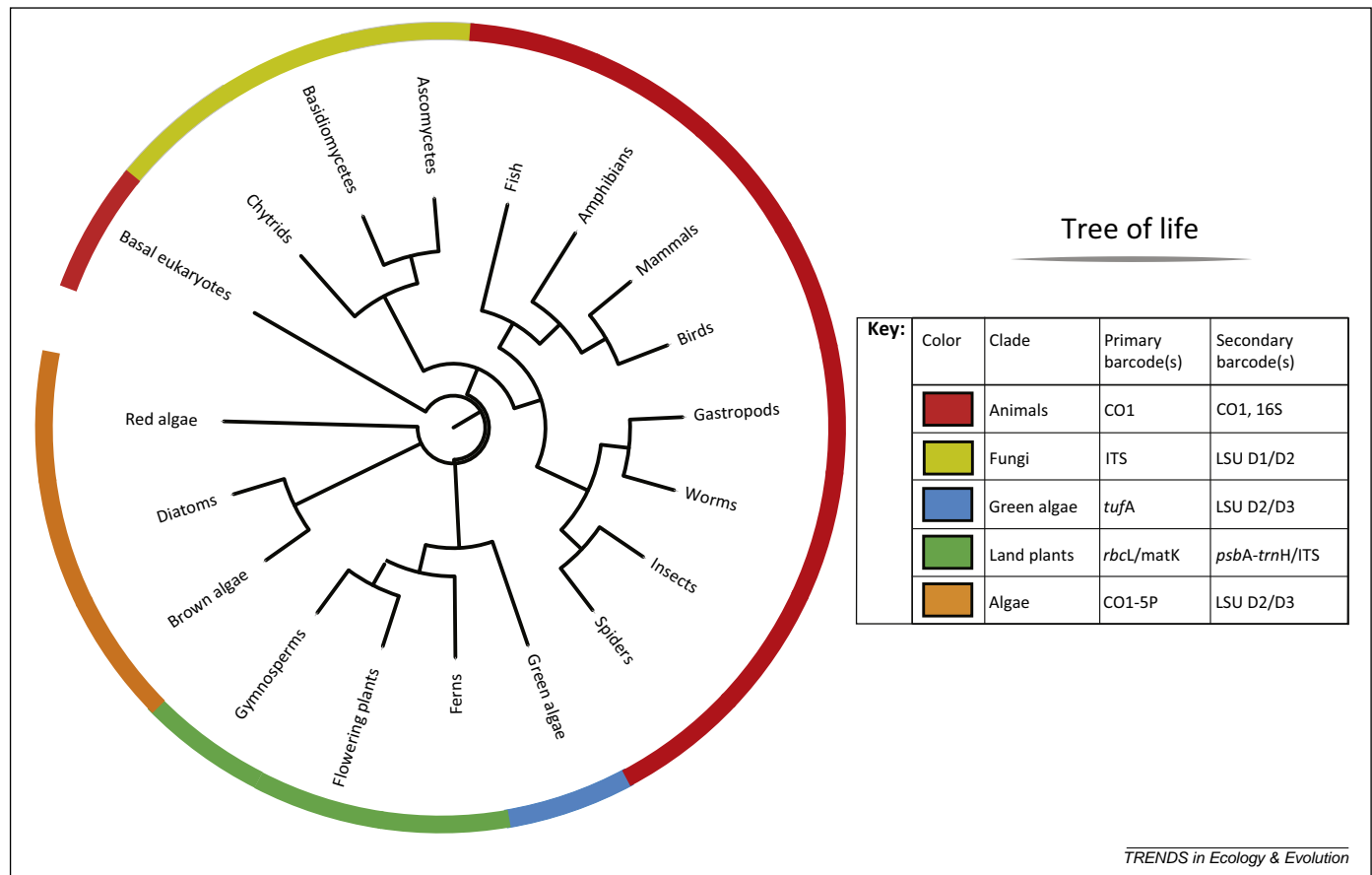


Figure 1. DNA barcodes are becoming a standard tool to discover, describe, and understand biodiversity, specifically to identify species in understudied, microscopic, and highly diverse lineages. DNA barcodes in conjunction with morphological, biochemical, and ecological information are revealing an outstanding diversity of species previously unrecognized through the analysis of morphological variation alone. During the first 8 months of 2014, the Web of Science recorded 310 publications in which DNA barcoding was used in discovering and describing new species, including algae [78], ferns [79], fungi [80], nematodes [81,82], arthropods [83,84], mollusks [85], fish [25], birds [86], and mammals [87]. Early on, it was expected that a single genetic marker would serve as a universal DNA barcode to identify species across the eukaryotic Tree of Life. In practice, different regions of DNA are necessary to provide adequate species identification across lineages. The plastid and nuclear loci indicated here are the most broadly used DNA barcodes for major groups of organisms, including algae [88], land plants [4,89], fungi [5], invertebrates [6], amphibians [7], fish [90], birds [70], and mammals [87,91]. Abbreviations: COI, cytochrome oxidase I; ITS, internal transcribed spacer.

species should be genetically parallel, the variation in success of DNA barcodes across lineages suggests that rates of evolution and the processes of speciation are not uniform.

The primary application of DNA barcodes will continue to be the identification of unknown samples. With the adoption of NGS technologies in many DNA barcode investigations (e.g., metagenomic applications [13]), uses are already being expanded to answer both applied and basic biological questions. Even if DNA barcodes are not uniformly successful for unambiguous identifications across the entire Tree of Life, ecologists, evolutionary biologists, and conservationists are already adopting DNA barcodes as a tool in their respective fields (e.g., [14,15]).

Current evolutionary, ecological, and conservation research with DNA barcodes

Taxonomy, systematics, and species discovery

A primary goal of evolutionary biologists and ecologists is to understand the origin of species and the factors causing the disparity in species richness in different biomes across the globe. In many cases, the full diversity of species in a given region is still unknown, especially in the most biodiverse habitats [16]. DNA barcodes have

been particularly useful in the discovery of cryptic and previously unrecognized species of animals [17]. For insects, it has been demonstrated that new species can be revealed through a combination of ecological field observations and DNA barcode markers [18,19]. For example, a single species of common skipper butterfly found throughout Central America defined by morphological features of the adults in fact comprises numerous species that are clearly delineated by DNA barcode sequences in congruence with diets and features of the larvae [20] (Box 2). Similarly, cryptic species of hispine beetles and larval stages that were identified through a DNA barcode survey of plant–herbivore interactions in Costa Rica linked the adults with the larvae found on the host plants [21] (Box 2). An extensive DNA barcode study on Microgastrinae wasps demonstrated significant insights into their taxonomy and species discovery [22]. Such cryptic diversity has been uncovered using DNA barcodes in other animal taxa, including crustaceans, diatoms, and fish [23–25]. The use of DNA barcodes for the discovery of new species is emerging as a powerful tool to clarify species boundaries and to quantify species diversity [26]. In many cases, these genetic markers serve as the starting point for the discovery of new taxa (Figure 1).

Box 2. Discovering new species, cryptic life stages, and ecological interactions in natural populations

DNA barcodes are a valuable genetic tool to reveal cryptic species previously unrecognized through the analysis of standard morphological variation. When combined with ecological data, the DNA barcodes provide additional evidence for determining species boundaries (Figure I). The *Astraptes fulgerator* species complex is a classic example of the use of DNA barcodes as a tool to discover cryptic biodiversity. In the Guanacaste Conservation Area (Costa Rica), DNA barcodes, combined with host plant records and larval morphology, demonstrated that one species of skipper (*A. fulgerator*), was in fact a complex of ten different species [20].

Another challenge for taxonomists is to associate immature life stages with their adult forms (Figure II). It is becoming routine to use DNA barcodes to easily associate different life-history stages with ontogenetically different morphologies, such as eggs, larvae, and adults, in insects [92]. In this example from a tropical premontane forest in Costa Rica, a comprehensive DNA barcode library containing sequences of all species from a community of *Cephaloleia* and *Chelobasis* beetles was assembled from tissue of adults. This reference library was then used to quickly identify the species of immature stages.

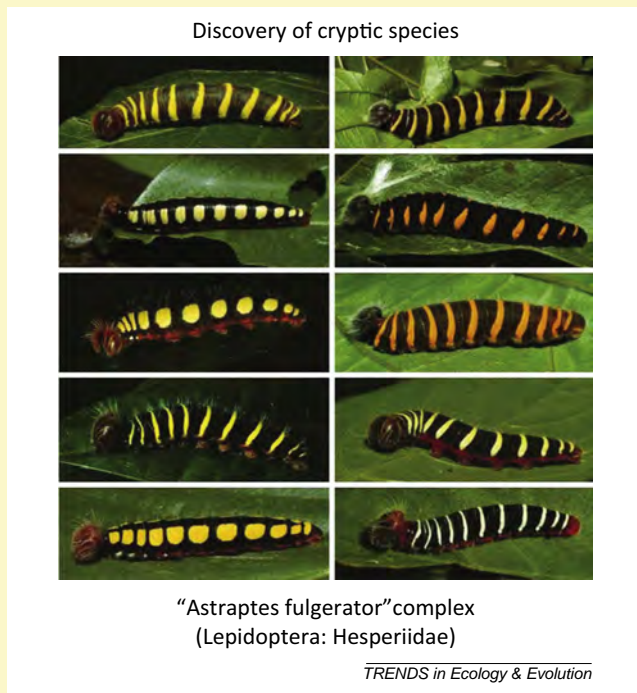


Figure I. DNA barcodes provide evidence to define species boundaries. Reproduced from [20].

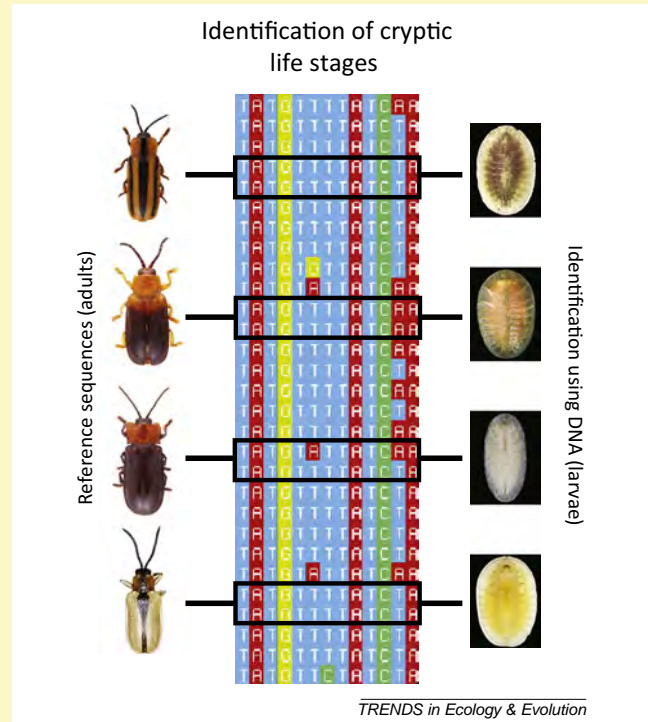


Figure II. DNA barcodes link life-history stages of insect species. Reproduced from [21].

Likewise, the potential exists for new plant species to be discovered and described as a result of genetic inventories based on both plastid and nuclear DNA barcode markers. For example, in the complex tropical plant family Lauraceae, the community phylogeny generated for the tree species on Barro Colorado Island with DNA barcode sequence data supported the recognition of a previously undescribed, but suspected, new species of *Nectandra* [27]. Furthermore, an ongoing DNA barcode survey of the trees in a forest dynamics plot located in the heart of the Amazon near Manaus, Brazil, suggested that many of the ‘morphospecies’ recognized by local taxonomists that do not currently have scientific names may have congruent support from the DNA barcode sequence data (Kress *et al.*, unpublished, 2014). This hyperdiverse research plot in the Amazon with over 1400 species of trees will be a test case for the utility of DNA barcodes in identifying potentially new species in a poorly known flora.

Community ecology and phylogeny

Over a decade ago, Webb [28] introduced a new set of methods and metrics to evaluate processes affecting community assembly in a phylogenetic context. This work demonstrated how phylogenetic relations among species

within a community can be analyzed to infer processes such as competition, environmental filtering, and trait evolution. Subsequent studies have been fruitful in mining phylogenetically structured community data to investigate alpha and beta species diversity [29,30], the linkage between phylogenetic diversity and dispersion [31,32], and the role of functional traits and evolutionary history in structuring communities [33,34] (Figures 2 and 3). One constraint on the use of phylogenetic data in community ecology has been the availability of phylogenies that accurately reflected the evolutionary relations among community members, particularly at lower taxonomic scales. Given that standard phylogenetic analyses are taxon driven, inclusion of all members of a specific community is rare. Thus, the availability of highly resolved phylogenetic trees that contain all members of a community has been critically lacking. The ability of DNA barcode data to help reconstruct evolutionary relations within targeted communities is helping to resolve this problem [27].

The first application of DNA barcodes in plant community phylogenetics was the reconstruction of the relationships of 281 tree species found in the Barro Colorado Island Forest Dynamics Plot in Panama [27]. That study demonstrated how the multilocus plant DNA barcode could robustly

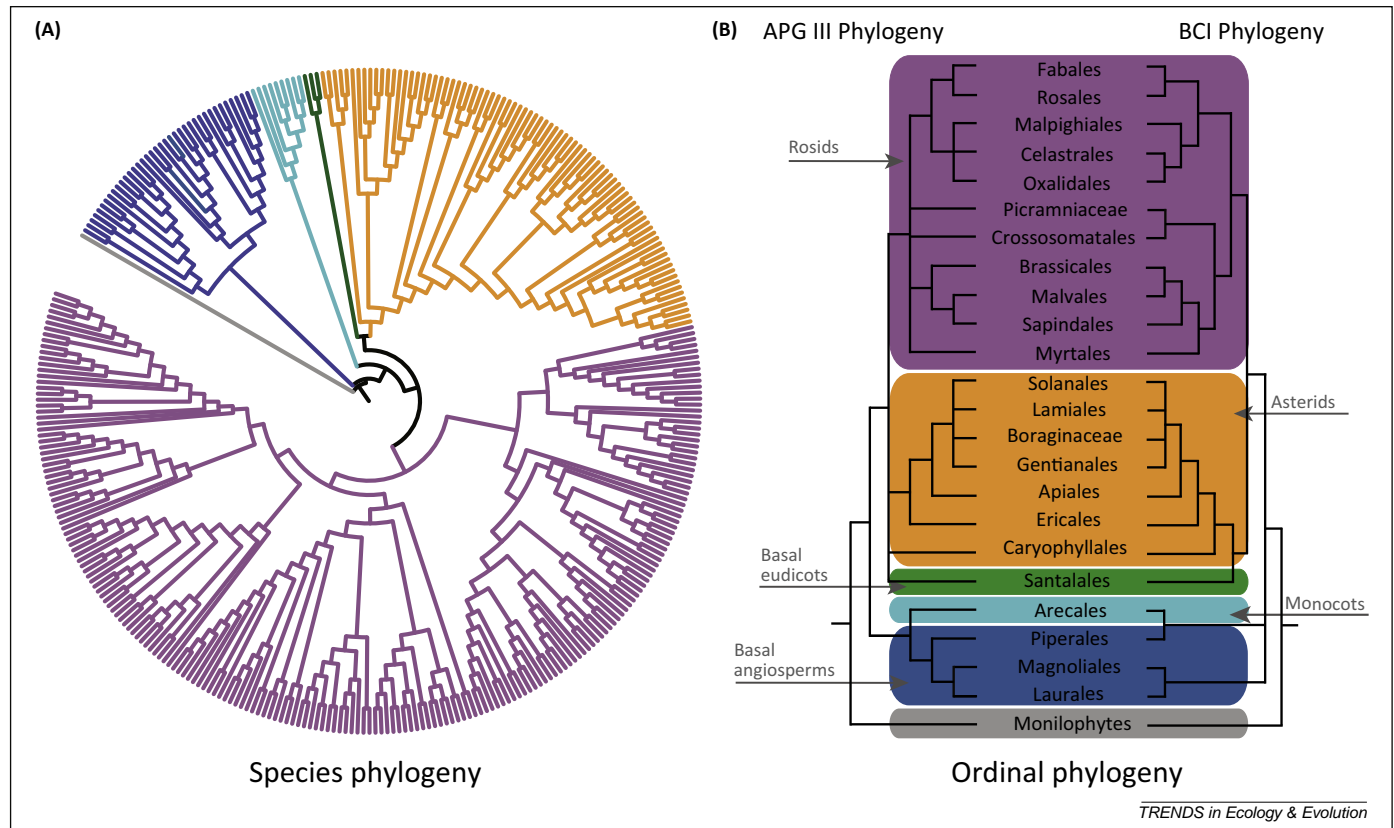


Figure 2. Sequence data from DNA barcodes can be used to reconstruct phylogenetic relations among the species of a community, especially in plants. The phylogeny resulting from the analyses of DNA barcode sequence data from 281 species of trees in the forest dynamics plot on Barro Colorado Island, Panama (A) was highly congruent with the broadly accepted phylogeny of flowering plants (B) and provided exceptional levels of species resolution at the tips of the phylogenetic branches. The phylogeny was estimated through alignment of all *rbcL* and *matK* sequences using back translation; at the same time, the highly variable *trnH-psbA* locus was aligned within families and then each alignment block was concatenated to the globally aligned *rbcL* and *matK*. These alignments produced a highly sparse supermatrix, with over 95% missing data. However, the globally aligned *rbcL* and *matK* allowed for estimation of deeper nodes, while the *trnH-psbA* marker was able to better resolve the tips of the tree. The phylogeny was constructed using maximum likelihood with the program GARLI. This highly resolved phylogeny became the keystone for several studies of community assembly in this tropical forest [93,94]. Reproduced from [27].

reconstruct evolutionary relations among phylogenetically disparate community members (Figure 2). The Panama study also showed that a DNA barcode-based community phylogeny has greatly improved topological resolution relative to super-tree methods that had been adopted previously. The improvement of topological resolution through incorporation of DNA barcode data for all or most species in the community improves the power of evolutionary-based hypothesis testing in an ecological framework. Since that first DNA barcode-based community phylogeny, similar investigations have focused on other forests in the tropics [27,35].

A major challenge of reconstructing a phylogenetic tree using DNA barcode sequence data is to capture the proper evolutionary relations among both highly divergent as well as closely related species. Given the DNA barcode requirement for relative short genetic markers (less than 1500 bp in total for plants), resolving phylogenetic relations among the basal branches and the tips of the branches is difficult. In this sense, one can regard these community phylogenies as a small part of a Tree of Life project, which seeks to assemble and reconstruct species relations. It has been shown that, despite limited nucleotide content, taxon relations can be well estimated when species density is high [36]. As more DNA barcode data are assembled, an increased number of taxa will be woven into increasingly larger phylogenies. This approach results in

more well-supported phylogenetic reconstructions of the constituent species from which individual community phylogenies may be pruned out for targeted analysis [35]. Furthermore, the use of a constraint tree derived from existing Tree of Life studies will enforce topological relations at deeper phylogenetic levels, where the limited signal from the small amount of nucleotide data provided by the DNA barcode markers is most problematic (Figure 3) [27]. The DNA barcode data are then relied upon to primarily resolve generic- and species-level relations towards the tips of the evolutionary branches.

Species assembly and functional trait evolution

The rapid increase in phylogenetic information based on DNA barcode sequence data [27] coupled with advances in functional trait-based community ecology [37,38] have been used to address questions of community assembly and diversity [39,40]. This merging of evolutionary and ecological perspectives has led to new insights into the role of ecological, biogeographical, and evolutionary processes in the distribution of biodiversity [41] (Figure 3).

Despite these advances, several conceptual and methodological challenges remain. One of the most salient issues is the difficulty of dynamically linking biophysical processes and species interactions that unfold over vastly different spatial and temporal scales (R. Muscarella *et al.*, unpublished, 2014). Most empirical studies that have

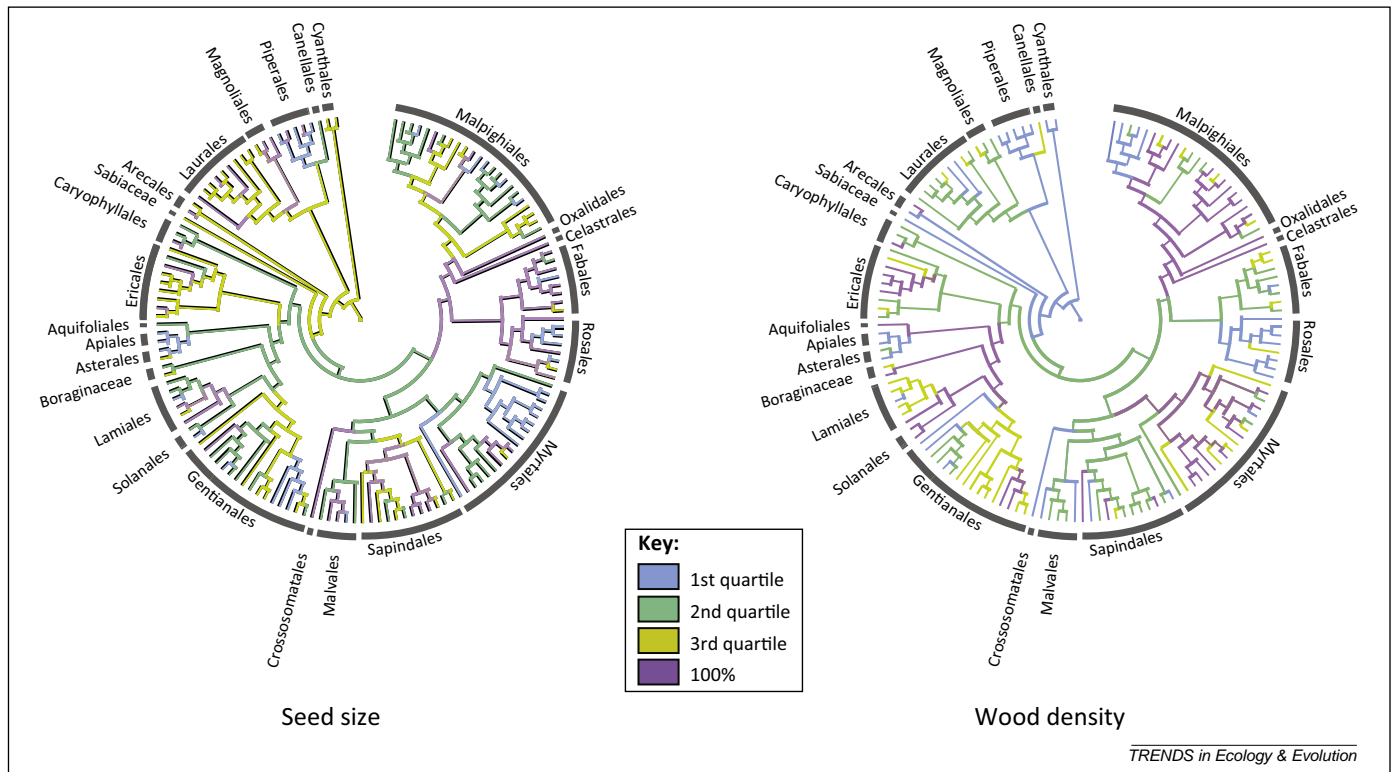


Figure 3. Well-resolved phylogenies derived from DNA barcodes coupled with functional trait data of the same species provide new insights into the processes of community assembly. In the Luquillo Forest Dynamics Plot in Puerto Rico, phylogenetic analyses based on DNA barcode sequence data [95] used a constraint tree to reconstruct the evolutionary history of the trees in this tropical forest community. A constraint tree derived from APGIII was implemented to take advantage of the prior knowledge and data used to estimate familial relations [96]; the phylogenetic constraint comprised the APG III phylogeny applied to all taxa in the study, but with the constraint truncated to a polytomy at the ordinal level. Therefore, the DNA barcode data only needed to resolve relations within each taxonomic order. The phylogenetic data combined with information on functional traits mapped onto the community phylogeny (such as seed size and wood density) helped to elucidate the importance of several community assembly processes (e.g., niche partitioning and competitive hierarchies) in this forest (e.g., [32,34]). several powerful programs are available within the R programming language that facilitate the joint analysis of phylogenetic and trait data (e.g., Picante; [97]). Reproduced from [34].

simultaneously investigated phylogenetic and functional community structure as a means to provide insights into community assembly processes have been focused at either the local (e.g., [29]) or global scales (e.g., [42]) rather than at intermediate (i.e., regional) scales, where local and global factors interact. Global-scale studies have examined the role of historical biogeography on community assembly without much attention to ecological processes. By contrast, local-scale studies have generally focused on disentangling the influence of habitat filtering and niche differentiation on community composition (e.g., [29]) or demographic rates (e.g., [34]). In most of these local studies, the use of phylogenies has been ‘corrective’, that is, largely limited to quantifying how the assumption of phylogenetic dependence between species [43] might influence results.

A greater emphasis on intermediate scales (i.e., beta diversity) that reflect how historical and local processes unfold over environmental gradients in space is likely to yield deeper insights into the role of evolutionary processes on community assembly. Recent efforts are moving in this direction (e.g., [44]). Several approaches have been introduced to characterize the relations between particular traits, evolutionary lineages, and variation in environmental conditions across sites (e.g., [44–46]). For instance, Pavoine [47] developed a novel method to identify the association of trait states and phylogenetic trees with

spatially variable environmental factors. At coarse spatial scales, the association of phylogenies with space, but not with environmental factors indicates that historical processes (e.g., colonization) predominate over environmental factors in explaining community composition. Advancing our understanding of the processes that structure beta diversity will require data on species performance and community composition collected across environmental gradients (e.g., soil or precipitation) and at multiple spatial and temporal scales. The relative ease of generating universal DNA barcode data for plants in a community will facilitate these broad spatial comparisons and is currently being carried out across forest dynamics plots around the globe [35].

Diet analyses, trophic interactions, and ecological forensics

DNA barcodes represent a unique opportunity to understand trophic interactions among organisms, especially in habitats that are difficult to access, such as the forest canopy or underground zones. On Barro Colorado Island, a tropical rain forest in Panama, DNA barcodes were used to identify the species of roots collected from soil cores [48]. By combining DNA barcode-based root identifications with spatial records of individual trees from a mapped vegetation plot, the relation between belowground and aboveground diversity was assessed in a diverse plant

community [48]. A similar study among plant species in a Canadian grassland also demonstrated success in reconstructing root interactions among species [49].

DNA barcodes can also assist in determining the diets of invertebrates, frugivorous birds, and elusive large mammals in the wild [50,51]. The first step in this DNA barcode approach to diet analyses is the collection of DNA tissue samples of the prey or host taxa under investigation (Box 1). For invertebrate herbivores, once the library is complete, DNA of host species is isolated directly from the gut contents of the animals [52]. By contrast, analyses of vertebrate diets are not invasive and are based on DNA extracted from scats [53,54]. The barcode markers utilized for amplification will depend on the diet of the organism [55]. The standard DNA barcode loci used to investigate herbivore diets differ markedly in their ability to identify plants at different taxonomic levels because of varying base-pair substitution rates across taxa. The plant DNA barcode markers *rbcL* and *trn-L-intron*, although easy to amplify from gut contents and scats, are usually only useful to identify plant tissues to the level of taxonomic family or genus (Figure 1) [52]. The DNA barcode loci *ITS2*, *matK*, and the noncoding intergenic spacer *trnH-psbA* are able to identify diet items to genus and species [52]. Unfortunately, the amplification of these DNA regions can be challenging or even impossible for some plant groups [52]. For animal prey, the most broadly used DNA barcode marker to identify diets is mitochondrial COI (Figure 1) and when employed with host-specific blocking primers, can enable identification of extensive prey diversity [56].

DNA sequences recovered from gut contents and scats are usually short (~100–400 bp) and of low quality as a consequence of DNA degradation during digestion [57,58]. The implementation of extraction techniques originally developed for ancient and antique DNA may improve the quality of the resultant DNA sequences [52]. However, DNA degradation during digestion is still a major limiting factor in diet analyses. When the DNA recovered is degraded, a mini-barcode version of the full DNA barcode marker [52] or a DNA barcode with more forensic applications [59] is required, both of which can not only improve amplification success, but also reduce rates of species-level identification.

The DNA sequencing approach to be used depends on the diet breadth of the consumer. For species that feed on one or only a few species during each feeding bout, it is possible to use traditional Sanger sequencing techniques (Box 1). For polyphagous species in which all diet items are more difficult to identify, it is now possible to accurately determine all consumed species using NGS methodology [60] (Box 3).

The most efficient method to identify the taxa consumed by a particular animal or group of animals is to compare DNA sequences extracted from gut contents and scats directly to a reference DNA barcode library of the food items (Box 4). The quality of identifications will depend on the completeness and the accuracy of the sequences included in the reference library. Researchers should be cautious when using public DNA sequence repositories, such as GenBank, because of their incompleteness. When DNA sequences for potential food items are missing in the

database, the contents of guts and scats can be misidentified. The error rate may also increase because of faulty taxonomic identifications and lack of verifiable taxonomic voucher specimens [57]. However, the scope of these taxonomic errors in open-source databases has not been quantified.

A solution proposed by recent studies is to generate *a priori* a comprehensive and well-curated DNA barcode library of potential diet items [21,50,52]. By generating complete DNA reference libraries for food plants, diet identifications can be accurate, even to the species level, and the error rate will be reduced or even eliminated [52] (Box 4). As NGS technologies become more accessible and cost-effective, more researchers will rely on these techniques to investigate and understand the complexity of trophic interactions in nature.

Conservation biology: quantifying species richness and evolutionary diversity within and among communities

The level of biological diversity present in an environment can be quantified by either enumerating numbers of species (e.g., Simpson's diversity) or estimated evolutionary divergences among species in which genetic distances have been calculated [28]. Although most measures of alpha and beta diversity across plant communities are based on numbers of species, DNA sequence data, if available, will provide an evolutionary dimension to diversity estimates that incorporate genetic distance among species (i.e., phylogenetic branch lengths). Evolutionary diversity, also called phylogenetic diversity [61,62], is correlated with, but not equivalent to, species richness. For example, geographic areas that harbor relatively few species may have high phylogenetic diversity if the species present are broadly dispersed across the Tree of Life. It has also been shown that a discontinuity between species richness and evolutionary diversity occurs when there are concentrations of closely related species in an environment [63].

DNA barcodes can provide a universal marker across species in a community or a region by which genetic distance, hence phylogenetic diversity, can be quantified within and across ecological communities at varying geographic scales [64]. When compared with species richness in the same communities, these genetic measures can also be used to evaluate species boundaries, can serve as clues to assist in documenting new species, and can identify targeted habitats for conservation [61,65]. In geographic regions known for their especially unique lineages of plants and animals, such as northeast Queensland in Australia and South Africa, phylogenetic diversity defined with DNA barcode sequence data may be the most important measure for comparing diversity and establishing protected areas across the landscape (A. Shapcott *et al.*, unpublished, 2014). As the DNA barcode library becomes populated with species across the globe, comparative measures of phylogenetic diversity will become standard metrics for conservation assessment.

In addition to the assessment of biodiversity hotspots for conservation, DNA barcodes are now also being used for the reliable identification and detection of illegally traded and often endangered species [66]. Similarly, genetic identifications using DNA barcodes for wild-collected medicinal

Review

plants [67], market adulterants in certified natural and commercial products [68], and genetically modified crops [69] are becoming more common. These applied uses of DNA barcodes for conservation and commercial purposes will undoubtedly increase in the future, especially as sequencing technology becomes simpler and less expensive.

Concluding remarks and future contributions of DNA barcodes

Here, we have focused mainly on the basic scientific applications of DNA barcodes to increase our understanding of species relations and boundaries, community ecological processes and networks, and the assessment of biodiversity for effective conservation. In addition, the forensic use of DNA barcodes for identification of endangered species and commercially useful plants and animals is being expanded by local, state, and national governments. DNA barcodes are proving useful as evidence in criminal cases and investigations of natural and manmade disasters. For example, a library of CO1 markers for birds is now routinely used to identify avian species involved in airplane strikes [70]. These applications are only in their infancy, but may eventually be a major source of data for the growing DNA barcode library.

The most significant advance in DNA barcoding in the near future will be the application of new technologies for generating and analyzing DNA barcode sequences. Several studies [71,72] and reviews [72,73] have addressed the marriage of DNA barcoding and NGS, especially with regards to environmental sampling. In general, NGS platforms are not used for collecting and constructing DNA barcode reference libraries for a specific set of taxa. Rather, NGS will enable the capture of all representative sequences present in a complex mixture of species and

Box 3. DNA metabarcoding and environmental samples

Complex environmental samples containing multiple organisms now can be identified using metabarcoding (Figure 1). These techniques are based on NGS technologies. DNA sequences from complex mixtures of organisms representing different species are obtained through NGS that localizes and simultaneously recovers sequence data from all individuals in a sample. A major challenge in DNA metabarcoding analyses is to translate the resulting millions of sequences into a list of taxonomic units. Several bioinformatics tools are now available to identify taxa from these large data sets. One option is to identify taxa from these samples by grouping DNA sequences into molecular operational taxonomy units (MOTUs) and then comparing consensus sequences from each MOTU to a reference library.

For DNA metabarcoding, many of the methods rely on PCR coupled with NGS to recover those sequences that were successfully amplified from the mixture. This reliance on PCR points to one of the two main challenges associated with DNA meta-barcoding when applied to complex mixtures of sequences. PCR bias in amplification may result in skewed recovery of sequences from the environmental mixture [98]. In this case, the problem is that the large amount of data generated might not include sequences that are representative of all species in the pool of organisms in the original mixture [99]. The second challenge is sequencing mistakes that are due to the high error rates in both PCR and NGS platforms. These errors can in turn affect assignment of sequences to the correct species in a reference database. However, solutions to both of these problems are likely in the near future through both advancements in bioinformatics and improved molecular techniques.

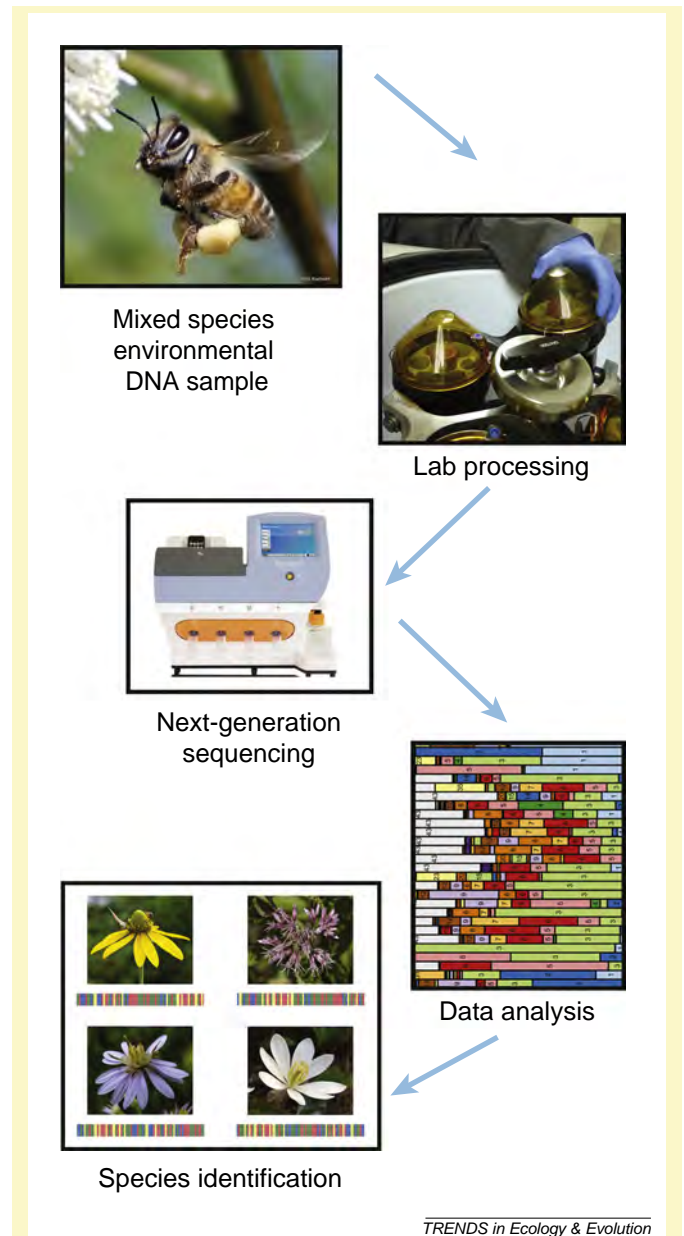


Figure 1. Workflow for using DNA metabarcodes and NGS for identifying environmental samples.

then the mapping of those sequences to a reference DNA barcode database [72]. These complex mixtures can be environmental samples of water or soil used for biodiversity assessment [74], or the mixtures can be targeted samples, such as animal scats or pollen loads on pollinators, for examining diet choice or pollinator foraging behavior (Box 3) [73,75]. This application of DNA barcoding to identify component species of an environmental mixture is termed 'DNA metabarcoding' [75]. As is true in other fields, the application of NGS to DNA barcoding will lead to tremendous growth in available sequence data, which themselves will require new tools for analysis as well as new systems for information storage.

Over the past 10 years, DNA barcoding has become an invaluable addition to our suite of tools to better understand nature and the environment. As the DNA barcode library expands across the Tree of Life, habitats, and

Box 4. Understanding ecological interactions in complex networks of species

Novel molecular techniques based on DNA barcodes can reconstruct interactions in complex networks of species in communities, including antagonistic and mutualistic plant-animal interactions. At La Selva Biological Station (Costa Rica), interactions between plants in the order Zingiberales and associated herbivores (chrysomelid rolled-leaf beetles) were determined by combining field records and DNA barcode analyses (Figure I). Here, each black rectangle represents a

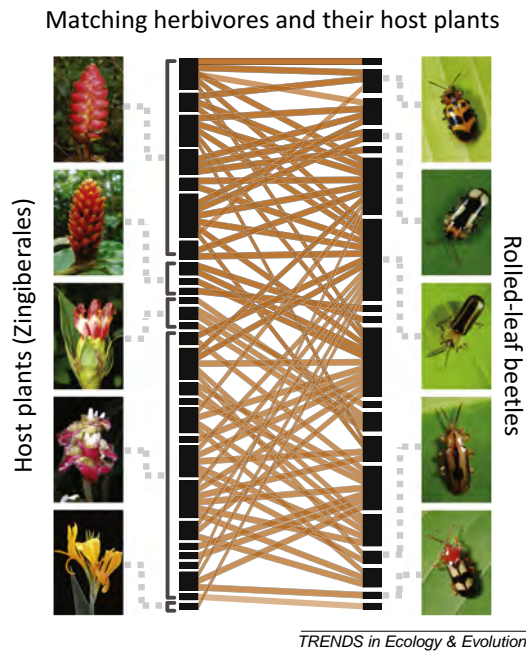


Figure I. DNA barcodes enable matching of host plants to their herbivores. Reproduced from [52].

plant or insect herbivore species. Lines represent single and multiple plant–herbivore associations. These molecular techniques have the potential to become a standard methodology for a detailed understanding of plant–herbivore interactions.

At the Garrapilos Mediterranean lowland forest in Spain, seed dispersal by birds was determined using a novel DNA barcode technique (Figure II). After assembling a comprehensive DNA barcode library (using COI), mini-barcodes were developed to amplify degraded DNA for this particular bird assemblage. The bird species responsible for dispersing seeds were then identified by amplifying bird DNA obtained from the defecated propagules. The identification of interacting species represents a unique opportunity to understand the process of fruit selection by frugivores and the contribution of each frugivorous species to seed dispersal to different microsites [50]. These techniques have the potential to connect frugivory with seed dispersal in ways previously unattainable through traditional field techniques.

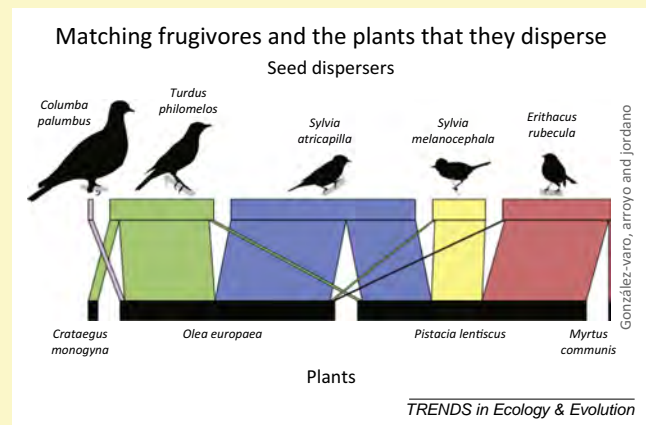


Figure II. DNA barcodes enable matching of seeds and fruits to their bird dispersers. Reproduced from [50].

geographies, new and more expansive applications will be explored and developed. We are only at the beginning of applying DNA barcodes to the fields of species discovery, ecology, evolution, and the conservation of biodiversity.

Acknowledgments

We are grateful to Susanne Renner and two anonymous reviewers of this manuscript for their constructive suggestions. Authors were supported by grants from the NSF-RCN program for ‘Tropical Forests in a Changing World’ (Award ID: 0741956) to D. McClearn, the Smithsonian Institution Postdoctoral Fellowship Program, Global Earth Observatories Program, and the Office of the Under Secretary for Science, and the National Geographic/Waitt Institute Grant (W149-11) to C.G.-R. M.U. acknowledges support from NSF DEB-0620910 to the Luquillo Long-Term Ecological Research Program. The assistance of I. Lopez in constructing the illustrations was invaluable. We would like to thank D. Janzen, N. Swenson, P. Hollingsworth, L. Weigt, A. Driscoll, P. Taberlet, J. Chave, C. Dick, C. Meyer, D. Schindel, and many others for helpful discussion on methods and uses of DNA barcodes.

References

- 1 Hebert, P.D.N. *et al.* (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. B* 270, S96–S99
- 2 Pecnikar, Z.F. and Buzan, E.V. (2014) 20 years since the introduction of DNA barcoding: from theory to application. *J. Appl. Genet.* 55, 4–52
- 3 Hebert, P.D.N. and Barrett, R.D.H. (2005) Reply to the comment by L. Prendini on ‘Identifying spiders through DNA barcodes’. *Can. J. Zool.* 83, 505–506
- 4 Kress, W.J. and Erickson, D.L. (2007) A Two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE* 2, e508
- 5 Eberhardt, U. (2012) Methods for DNA barcoding of fungi. In *DNA Barcodes: Methods and Protocols: Methods in Molecular Biology* (Kress, W.J. and Erickson, D.L., eds), pp. 183–206, Humana Press, Springer Science+Publishing Media, LLC
- 6 Evans, N. and Paulay, G. (2012) DNA barcoding methods for invertebrates. In *DNA Barcodes: Methods and Protocols. Methods in Molecular Biology* (Kress, W.J. and Erickson, D.L., eds), pp. 47–78, Humana Press, Springer Science+Publishing Media, LLC
- 7 Vences, M. *et al.* (2012) DNA barcoding amphibians and reptiles. In *DNA Barcodes: Methods and Protocols. Methods in Molecular Biology* (Kress, W.J. and Erickson, D.L., eds), pp. 79–108, Humana Press, Springer Science+Publishing Media, LLC
- 8 Spooner, D.M. (2009) DNA barcoding will frequently fail in complicated groups: an example in wild potatoes. *Am. J. Bot.* 96, 1177–1189
- 9 Elias, M. *et al.* (2007) Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proc. R. Soc. B* 274, 2881–2889
- 10 Fazekas, A.J. *et al.* (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Mol. Ecol. Resour.* 9, 130–139
- 11 Moritz, C. and Cicero, C. (2004) DNA barcoding: promise and pitfalls. *PLoS Biol.* 2, e354
- 12 Razafimandimbison, S.G. *et al.* (2004) Recent origin and phylogenetic utility of divergent ITS putative pseudogenes: a case study from Naucleaceae (Rubiaceae). *Syst. Biol.* 53, 177–192
- 13 Kress, W.J. and Erickson, D.L. (2008) DNA barcodes: genes, genomics, and bioinformatics. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2761–2762
- 14 Joly, S. *et al.* (2014) Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. *Mol. Ecol. Resour.* 14, 221–232
- 15 Chakraborty, C. *et al.* (2014) DNA barcoding to map the microbial communities: current advances and future directions. *Appl. Microbiol. Biotechnol.* 98, 3425–3436
- 16 Purvis, A. and Hector, A. (2000) Getting the measure of biodiversity. *Nature* 405, 212–219

- 17 Smith, M.A. *et al.* (2008) Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12359–12364
- 18 Shashank, P.R. *et al.* (2014) DNA barcoding reveals the occurrence of cryptic species in host-associated population of *Conogethes punctiferalis* (Lepidoptera: Crambidae). *Appl. Entomol. Zool.* 49, 283–295
- 19 Bertrand, C. *et al.* (2014) Mitochondrial and nuclear phylogenetic analysis with Sanger and next-generation sequencing shows that, in Area de Conservacion Guanacaste, northwestern Costa Rica, the skipper butterfly named *Urbanus belli* (family HesperIIDae) comprises three morphologically cryptic species. *BMC Evol. Biol.* 14, 153
- 20 Hebert, P.D.N. *et al.* (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc. Natl. Acad. Sci. U.S.A.* 101, 14812–14817
- 21 García-Robledo, C. *et al.* (2013) Using a comprehensive DNA barcode library to detect novel egg and larval host plant associations in *Cephaloleia* Rolled-leaf Beetles (Coleoptera: Chrysomelidae). *Biol. J. Linn. Soc.* 110, 189–198
- 22 Smith, M.A. *et al.* (2012) DNA barcoding and the taxonomy of Microgastrinae wasps (Hymenoptera, Braconidae): impacts after 8 years and nearly 20 000 sequences. *Mol. Ecol. Resour.* 13, 168–176
- 23 Silva, E.D. *et al.* (2014) *Alona iheringula* Sinev & Kotov, 2004 (Crustacea, Anomopoda, Chydoridae, Aloninae): life cycle and DNA barcode with implications for the taxonomy of the Aloninae subfamily. *PLoS ONE* 9, e97050
- 24 Hamsher, S.E. and Saunders, G.W. (2014) A floristic survey of marine tube-forming diatoms reveals unexpected diversity and extensive cohabitation among genetic lines of the *Berkeleya rutilans* complex (Bacillariophyceae). *Eur. J. Phycol.* 49, 47–59
- 25 Winterbottom, R. *et al.* (2014) A cornucopia of cryptic species: a DNA barcode analysis of the gobiid fish genus *Trimma* (Percomorpha, Gobiiformes). *Zookeys* 381, 79–111
- 26 Costion, C. *et al.* (2011) Plant DNA barcodes can accurately estimate species richness in poorly known floras. *PLoS ONE* 6, e26841
- 27 Kress, W.J. *et al.* (2009) Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18621–18626
- 28 Webb, C.O. (2000) Exploring the phylogenetic structure of ecological communities: an example for rain forest trees. *Am. Nat.* 156, 145–155
- 29 Kraft, N.J.B. *et al.* (2008) Functional traits and niche-based tree community assembly in an amazonian forest. *Science* 322, 580–582
- 30 Swenson, N.G. and Enquist, B.J. (2009) Opposing assembly mechanisms in a Neotropical dry forest: implications for phylogenetic and functional community ecology. *Ecology* 90, 2161–2170
- 31 Graham, C.H. and Fine, P.V.A. (2008) Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecol. Lett.* 11, 1265–1277
- 32 Swenson, N.G. *et al.* (2012) Phylogenetic and functional alpha and beta diversity in temperate and tropical tree communities. *Ecology* 93, S112–S125
- 33 Hardy, O.J. and Senterre, B. (2007) Characterizing the phylogenetic structure of communities by an additive partitioning of phylogenetic diversity. *J. Ecol.* 95, 493–506
- 34 Uriarte, M. *et al.* (2010) Trait similarity, shared ancestry and the structure of neighbourhood interactions in a subtropical wet forest: implications for community assembly. *Ecol. Lett.* 13, 1503–1514
- 35 Erickson, D.L. *et al.* (2014) Comparative evolutionary diversity and phylogenetic structure across multiple forest dynamics plots: a megaphylogeny approach. *Front. Genet.* Published online September 20, 2014. (<http://dx.doi.org/10.3389/fgene.2014.00358>)
- 36 Smith, M.A. *et al.* (2009) DNA barcode accumulation curves for understudied taxa and areas. *Mol. Ecol. Resour.* 9, 208–216
- 37 McGill, B.J. *et al.* (2006) Response to Kearney and Porter: both functional and community ecologists need to do more for each other. *Trends Ecol. Evol.* 21, 482–483
- 38 Westoby, M. and Wright, I.J. (2006) Land-plant ecology on the basis of functional traits. *Trends Ecol. Evol.* 21, 261–268
- 39 Webb, C.O. *et al.* (2002) Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* 33, 475–505
- 40 Cavender-Bares, J. *et al.* (2009) The merging of community ecology and phylogenetic biology. *Ecol. Lett.* 12, 693–715
- 41 Cavender-Bares, J. *et al.* (2004) Phylogenetic overdispersion in Floridian oak communities. *Am. Nat.* 163, 823–843
- 42 Qian, H. and Ricklefs, R.E. (2007) A latitudinal gradient in large-scale beta diversity for vascular plants in North America. *Ecol. Lett.* 10, 737–744
- 43 Harvey, P.H. and Pagel, M.D. (1991) *The Comparative Method in Evolutionary Biology*, Oxford University Press
- 44 Ives, A.R. and Helmus, M.R. (2011) Generalized linear mixed models for phylogenetic analyses of community structure. *Ecol. Monogr.* 81, 511–525
- 45 Mayfield, M.M. *et al.* (2009) Traits, habitats, and clades: identifying traits of potential importance to environmental filtering. *Am. Nat.* 174, E1–E22
- 46 Leibold, M.A. *et al.* (2010) Metacommunity phylogenetics: separating the roles of environmental filters and historical biogeography. *Ecol. Lett.* 13, 1290–1299
- 47 Pavoine, S. *et al.* (2010) Decomposition of trait diversity among the nodes of a phylogenetic tree. *Ecol. Monogr.* 80, 485–507
- 48 Jones, F.A. *et al.* (2011) The roots of diversity: below ground species richness and rooting distributions in a tropical forest revealed by DNA Barcodes and inverse modeling. *PLoS ONE* 6, e24506
- 49 Kesanakurti, P.R. *et al.* (2011) Spatial patterns of plant diversity below-ground as revealed by DNA barcoding. *Mol. Ecol.* 20, 1289–1302
- 50 Gonzalez-Varo, J.P. *et al.* (2014) Who dispersed the seeds? The use of DNA barcoding in frugivory and seed dispersal studies. *Methods Ecol. Evol.* 5, 806–814
- 51 Deagle, B.E. and Tollit, D.J. (2007) Quantitative analysis of prey DNA in pinniped faeces: potential to estimate diet composition? *Conserv. Genet.* 8, 743–747
- 52 Garcia-Robledo, C. *et al.* (2013) Tropical plant–herbivore networks: reconstructing species interactions using DNA barcodes. *PLoS ONE* 8, e52967
- 53 Shehzad, W. *et al.* (2012) Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Mol. Ecol.* 21, 1951–1965
- 54 Clare, E.L. *et al.* (2009) Species on the menu of a generalist predator, the eastern red bat (*Lasiurus borealis*): using a molecular approach to detect arthropod prey. *Mol. Ecol.* 18, 2532–2542
- 55 Hibert, F. *et al.* (2013) Unveiling the diet of elusive rainforest herbivores in Next Generation Sequencing era? The tapir as a case study. *PLoS ONE* 8, e60799
- 56 Leray, M. *et al.* (2013) Effectiveness of annealing blocking primers versus restriction enzymes for characterization of generalist diets: unexpected prey revealed in the gut contents of two coral reef fish species. *PLoS ONE* 8, e58076
- 57 Jurado-Rivera, J.A. *et al.* (2009) DNA barcoding insect-host plant associations. *Proc. R. Soc. B* 276, 639–648
- 58 Pinzón-Navarro, S. *et al.* (2010) DNA profiling of host-herbivore interactions in tropical forests. *Ecol. Entomol.* 35, 18–32
- 59 Epp, L.S. *et al.* (2012) New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Mol. Ecol.* 21, 1821–1833
- 60 Pompanon, F. *et al.* (2012) Who is eating what: diet assessment using next generation sequencing. *Mol. Ecol.* 21, 1931–1950
- 61 Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* 61, 1–10
- 62 Helmus, M.R. *et al.* (2007) Phylogenetic measures of biodiversity. *Am. Nat.* 169, E68–E83
- 63 Forest, F. *et al.* (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* 445, 757–760
- 64 Chen, S. *et al.* (2010) Validation of the *ITS2* region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* 5, e8613
- 65 Faith, D.P. (2008) Phylogenetic diversity and conservation. In *Conservation Biology: Evolution in Action* (Carroll, S.P. and Fox, C., eds), pp. 99–115, Oxford University Press
- 66 Lahaye, R. *et al.* (2008) A DNA barcode for the flora of the Kruger National Park (South Africa). *S. Afr. J. Bot.* 74, 370–371
- 67 Purushothaman, N. *et al.* (2014) A tiered barcode authentication tool to differentiate medicinal *Cassia* species in India. *Genet. Mol. Res.* 13, 2959–2968

- 68 Swetha, V.P. *et al.* (2014) DNA barcoding for discriminating the economically important *Cinnamomum verum* from its adulterants. *Food Biotechnol.* 28, 183–194
- 69 Nzeduru, C.V. *et al.* (2012) DNA barcoding simplifies environmental risk assessment of genetically modified crops in biodiverse regions. *PLoS ONE* 7, e35929
- 70 Dove, C.J. *et al.* (2008) Using DNA barcodes to identify bird species involved in birdstrikes. *J. Wildl. Manag.* 72, 1231–1236
- 71 Deagle, B.E. *et al.* (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Mol. Ecol.* 18, 2022–2038
- 72 Valentini, A. *et al.* (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Mol. Ecol. Resour.* 9, 51–60
- 73 Taberlet, P. *et al.* (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* 21, 2045–2050
- 74 Sønsteby, J.H. *et al.* (2010) Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Mol. Ecol. Resour.* 10, 1009–1018
- 75 Taberlet, P. *et al.* (2012) Environmental DNA. *Mol. Ecol.* 21, 1789–1793
- 76 Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425
- 77 Altschul, S.F. *et al.* (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410
- 78 Machin-Sanchez, M. *et al.* (2014) A combined barcode and morphological approach to the systematics and biogeography of *Laurencia pyramidalis* and *Laurenciella maritzae* (Rhodophyta). *Eur. J. Phycol.* 49, 115–127
- 79 Dauphin, B. *et al.* (2014) Molecular phylogenetics supports widespread cryptic species in Moonworts (*Botrychium* ss, Ophioglossaceae). *Am. J. Bot.* 101, 128–140
- 80 Yap, H.Y.Y. *et al.* (2014) DNA barcode markers for two new species of Tiger Milk Mushroom: *Lignosus tigris* and *L. cameronensis*. *Int. J. Agric. Biol.* 16, 841–844
- 81 Felix, M.A. *et al.* (2014) A streamlined system for species diagnosis in *Caenorhabditis* (Nematoda: Rhabditidae) with name designations for 15 distinct biological species. *PLoS ONE* 9, e94723
- 82 Powers, T.O. *et al.* (2014) COI haplotype groups in Mesocriconema (Nematoda: Criconematidae) and their morphospecies associations. *Zootaxa* 3827, 101–146
- 83 Pillai, P.M. *et al.* (2014) Description, DNA barcode and phylogeny of a new species, *Macrobrachium abrahami* (Decapoda: Palaemonidae) from Kerala, India. *Zootaxa* 3768, 546–556
- 84 Stefan, L.M. *et al.* (2014) A new species of the feather mite genus *Rhinozachvatkinia* (Acari: Avenzoariidae) from *Calonectris* shearwaters (Procellariiformes: Procellariidae): integrating morphological descriptions with DNA barcode data. *Folia Parasitol. (Praha)* 61, 90–96
- 85 Layton, K.K.S. *et al.* (2014) Patterns of DNA barcode variation in Canadian marine molluscs. *PLoS ONE* 9, e95003
- 86 Saitoh, T. *et al.* (2014) DNA barcoding reveals 24 distinct lineages as cryptic bird species candidates in and around the Japanese Archipelago. *Mol. Ecol. Resour.* Published online Jun 11, 2014. (<http://dx.doi.org/10.1111/1755-0998.12282>)
- 87 Wilson, J.J. *et al.* (2014) Utility of DNA barcoding for rapid and accurate assessment of bat diversity in Malaysia in the absence of formally described species. *Genet. Mol. Res.* 13, 920–925
- 88 Saunders, G.W. and McDevitt, D.C. (2012) Methods for DNA barcoding photosynthetic protists emphasizing the macroalgae and diatoms. In *DNA Barcodes: Methods and Protocols: Methods in Molecular Biology* (Kress, W.J. and Erickson, D.L., eds), pp. 207–222, Humana Press, Springer Science+Publishing Media, LLC
- 89 Hollingsworth, P.M. *et al.* (2009) A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12794–12797
- 90 Weigt, L.A. *et al.* (2012) DNA barcoding fishes. In *DNA Barcodes: Methods and Protocols: Methods in Molecular Biology* (Kress, W.J. and Erickson, D.L., eds), pp. 109–126, Humana Press, Springer Science+Publishing Media, LLC
- 91 Ivanova, N.V. *et al.* (2012) DNA barcoding in mammals. In *DNA Barcodes: Methods and Protocols: Methods in Molecular Biology* (Kress, W.J. and Erickson, D.L., eds), pp. 153–182, Humana Press, Springer Science+Publishing Media, LLC
- 92 Pinzón-Navarro, S. *et al.* (2010) DNA-based taxonomy of larval stages reveals huge unknown species diversity in neotropical seed weevils (genus *Conotrachelus*): relevance to evolutionary ecology. *Mol. Phylogenet. Evol.* 56, 281–293
- 93 Westbrook, J.W. *et al.* (2011) What makes a leaf tough? Patterns of correlated evolution between leaf toughness traits and demographic rates among 197 shade-tolerant woody species in a Neotropical forest. *Am. Nat.* 177, 800–811
- 94 Schreeg, L.A. *et al.* (2010) Phylogenetic analysis of local-scale tree soil associations in a lowland moist tropical forest. *PLoS ONE* 5, e13685
- 95 Kress, W.J. *et al.* (2010) Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican forest dynamics plot. *PLoS ONE* 5, e15409
- 96 Bremer, B. *et al.* (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121
- 97 Kembel, S.W. *et al.* (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26, 1463–1464
- 98 Thomas, T. *et al.* (2012) Metagenomics: a guide from sampling to data analysis. *Microb. Inform. Exp.* 2, 1–12
- 99 Deagle, B.E. *et al.* (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biol. Lett.* 10, 20140562