Chapter 18

Plant DNA Barcodes, Taxonomic Management, and Species Discovery in Tropical Forests

Christopher W. Dick and Campbell O. Webb

Abstract

DNA barcodes have great potential for species identification and taxonomic discovery in tropical forests. This use of DNA barcodes requires a reference DNA library of known taxa with which to match DNA from unidentified specimens. At an even more basic level, it presupposes that the species in the regional species pool have Latin binomials. This is not the case in species-rich tropical forests in which many species are new to science or members of poorly circumscribed species complexes. This chapter describes a workflow geared toward taxonomic discovery, which includes the discovery of new species, distribution records, and hybrid forms, and to management of taxonomic entities in forest inventory plots. It outlines the roles of laboratory technicians, field workers and herbarium-based taxonomists, and concludes with a discussion of potential multilocus nuclear DNA approaches for identifying species in recently evolved clades.

Key words: Tropical trees, Metadata, Vouchers, Taxonomy, Herbarium, DNA barcode, Discovery

1. Introduction

Tropical forests contain over 90% of the world's tree diversity (1). Single hectares of highly diverse Asian or South American forests contain more tree species than are found in the whole of eastern North America, or in the vast circumboreal forests. Yet our knowledge of these tropical plant species is poor, and the recorded species are only a fraction of the true species pool in most areas. For example, the *Catalog of the Vascular Plants of Ecuador* (2) lists only ca. 4,000 vascular plant species for Ecuadoran Amazon (a low intensity sampling of seven million hectares), whereas exhaustive sampling of a single hectare forest in the Ecuadorian Amazon yielded over 900 vascular plant species (3). Based on simple area versus species richness relationships, it is thus likely that the real tree richness of the region is many times 4,000 species (4), and that

W. John Kress and David L. Erickson (eds.), *DNA Barcodes: Methods and Protocols*, Methods in Molecular Biology, vol. 858,
DOI 10.1007/978-1-61779-591-6_18, © Springer Science+Business Media, LLC 2012

many of the additional species will be new to science; Prance (5) has estimated that 1 in 100 plant specimens collected from remote tropical forests, such as those in the Amazon basin and Papua New Guinea, come from as yet undescribed species.

Both the high species diversity and the high probability of encountering undescribed or at least poorly delineated taxa create challenges for biologists working on the inventory of tropical forest plants. One of the major challenges of setting up an inventory plot in tropical forest is simply keeping track of the morphotype identity of each tree, which must then be followed by the lengthy process of matching the morphotypes to known species. DNA barcodes can assist at both stages, accelerating the matching of morphotypes across a plot (via matching DNA haplotypes), and then matching taxonomic entities in the plot to named taxa.

Many ecologists involved with tropical forest inventory are excited by the prospect of using DNA barcodes to identify species (including new species), but they may not be familiar with (1) collection methods for molecular samples, (2) standard methods for making herbarium vouchers, or (3) the kinds of metadata that are needed to create DNA barcode reference libraries and describe new species. For a relatively low additional cost, an inventory program not primarily focused on plant taxonomy (e.g., inventory of carbon or the forest dynamics studies; ref. 6) can make high-quality DNA and physical vouchers which can be used to address urgent biodiversity questions.

1.1. DNA barcoding links to Systematics and Ecology

While traditional taxonomic work (e.g., collecting, matching, describing, publishing) will eventually increase global estimates of tropical tree diversity and yield new taxa, the specimens from botanical inventories often spend years in storage before they can be identified or described as new species (7). The "taxonomic bottleneck" is more pronounced now than ever due to the global decline in numbers of taxonomic specialists working on tropical groups. New species descriptions are also limited by inadequate representation of taxa in herbarium collections: herbarium-based specialists cannot describe the many new species that are not extremely distinctive because data pertaining to geographic range and morphological variation is frequently unavailable.

DNA barcodes can assist in this process of "taxonomic discovery," which can take the form of expanding species range information, delimiting species in closely related taxa, standardizing the nomenclature of species with multiple names (synonymy), or even recognizing species that are new to science. The effectiveness of DNA barcodes for taxonomic discovery depends on the preexistence of DNA records for many of the taxa under study, the so-called DNA reference library, but these libraries for tropical plant taxa are currently being built rapidly, primarily by ecologists working on forest inventories.

Tropical forest inventory plots provide several advantages for taxonomic discovery. Because the trees are tagged, the forest plots serve as living museums in which individual trees or their conspecific populations may be revisited to obtain additional data. Tropical forest plots delimit enough of the local flora to facilitate development of DNA barcode reference libraries for use in broader regional studies. Finally, the plot networks already have many of the human resources and institutional ties to universities and herbaria that are needed to sustain long-term research. An end goal of DNA-based forest inventory should be to standardize taxonomy across regional networks of forest plots, and thereby advance botanical knowledge of these relatively unexplored regions (6).

For example, to date, very few of the sets of vouchers from any Center for Tropical Forest Sciences (CTFS) Forest Dynamics Plot have been compared to a set from another plot because of the cost and logistical difficulty of such a cross-plot matching; DNA barcode matching among plots, on the other hand, could be done nearly instantaneously if the data were available.

The transfer of physical specimens and data between field workers, lab technicians, and systematists can be organized as a "workflow." Field workers collect the specimens and metadata, passing the plant tissue to a molecular laboratory and the pressed specimens to a herbarium (Fig. 1). The DNA barcode data then can assist both the field ecologists with basic management of taxonomic entities (morphotyping and matching vouchers), and the herbarium workers with matching to named species, and with range extensions, new species discovery, etc. There should be established lines of communication between laboratory technicians, systematists, and the field workers (or their local supervisor). For example, the lab workers may need additional plant tissue (e.g., from the vascular cambium instead of leaves) if PCR repeatedly fails. The plot workers should receive training from systematists prior to the collections, if possible, because taxonomic specialists often have tips for collecting taxonomically diagnostic field information for their particular groups (e.g., ref. 8).

The components of the total workflow that we focus on in this chapter are those geared toward the field ecologists, especially graduate students and postdocs, who work in forest inventory plots, such as the large (25–50 ha) tree inventory plots managed by CTFS (see Chapter 22) or the RAINFOR network of smaller (e.g., 1 ha) forest inventory plots scattered across the Amazon basin (9). We touch briefly on field methods, but make reference to additional sources. We do not emphasize molecular methods, which are described in detail by Fazekas et al. (Chapter 11) in this volume. We end with a discussion of multilocus nuclear DNA markers, which, in addition to the standard chloroplast DNA barcodes, will be necessary to rigorously test taxonomic hypotheses in forest inventory plots.

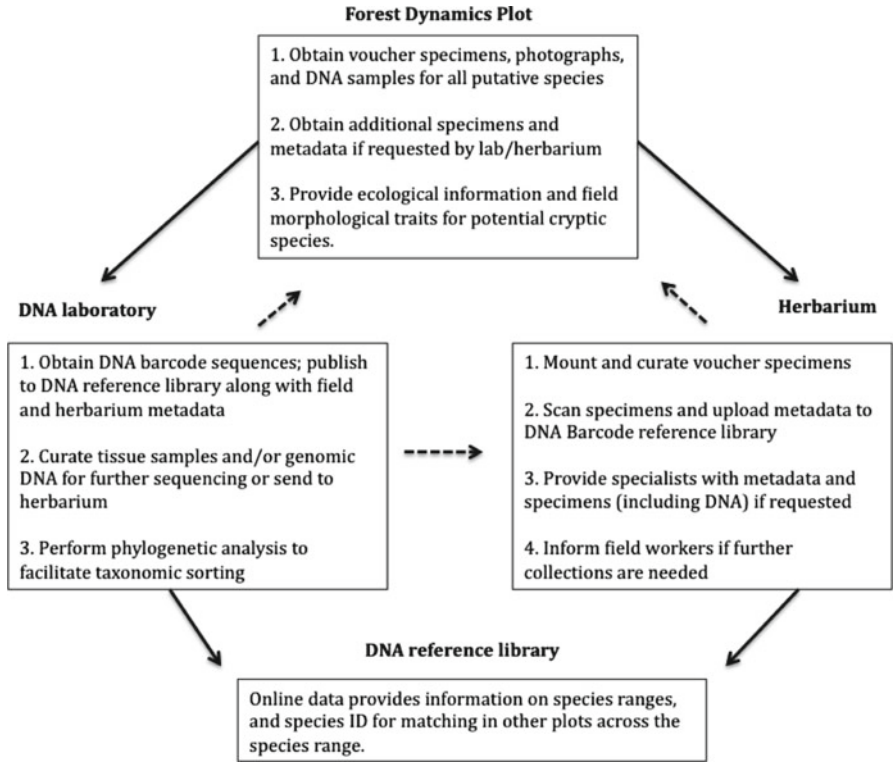

Fig. 1. Summary of the workflow that uses DNA barcodes for the purpose of taxonomic discovery in tropical forest inventory plots.

2. Field Materials

The materials for a DNA barcode project will depend on the magnitude of the project, for example, whether the goal is complete taxonomic inventory of a large forest plot, or collection and identification of focal groups by graduate students. The following materials would be especially useful for a graduate student getting involved in a taxonomic inventory project.

1. Collecting equipment: CWD uses a Jameson 2.43 m (8 ft) fiberglass pole clipper (Sherrill Inc.) with heavy-duty head for cutting branches up to 4.5 cm and four additional poles for a reach of 9.75 m (32 ft). Additional rope (approximately 40 ft) is needed with extensions. Additionally, local botanists need hand clippers and machetes with leather sheaths, wrist mounted slingshots and replacement rubber tubing, and climbing equipment to access the forest canopy. The kinds of climbing equipment vary from canvas belts used to shimmy up tree trunks to more elaborate single rope climbing techniques. The methods for describing rope-climbing techniques for tree climbing are beyond the scope of this chapter. Safety issues are a prime

concern, and should be considered and implemented, even if field workers are willing to take risks, and an insurance policy should be explicitly established for village assistants before they climb. Ref. 10 outlines safety considerations. A rubber mallet and >1 in. gasket hole punch may be needed to obtain vascular cambium tissue for DNA extraction (11).

2. Specimen drying equipment: A portable plant dryer can be made using a space heater and canvas cloth (12); and propane-fueled dryers with plywood frames are available in many large research stations. Plant presses for a drying oven should include the wooden mounting boards, tightening straps, blotting paper or cardboard, newspapers, and corrugated aluminum sheets to spread heat into the plant bundles (available at Forestry Suppliers Inc.). Enough newspapers should be available for layering between individual specimens.
3. Camera equipment: An optimal setup for making photographic vouchers would be a 35-mm digital camera with zoom lens and macro capacity, a ring flash for close-up shots, a tripod and black or gray cloth to use as standardized background. However, excellent results can be obtained with a high-end compact camera with built-in flash and macro (we recommend the Panasonic Lumix LX or GF series, or Canon G series). If in the field in the wet tropics for more than a few weeks, the camera and lens should be stored in an airtight container containing a desiccant, such as silica gel, as the humidity is conducive to the growth of fungi that destroy lens coatings.
4. Miscellaneous items: A large amount of relatively inexpensive equipment is needed to make field collections. For a more extensive list of items, refer to "Field Techniques used by Missouri Botanical Garden" (13). The items should include field notebooks that are small enough so as to contain relatively little information if lost (e.g., Rite in the Rain brand); garbage bags and Ziploc bags of varied sizes; alcohol and specimen jars to preserve small flowers; hand-held Geographic Positioning System (GPS) capable of fast reception under the forest canopy (e.g., Garmin CSx series), and fine silica gel with indicator beads to preserve leaf tissue for DNA extraction. Although fine silica gel is available from scientific supply companies, this high-grade variety is expensive and can be substituted with silica gel from florists.

3. Methods

3.1. Field Collections: What, When, and How

1. Replicate sampling: The collections for each morphospecies should include multiple individuals ($n = 3–5$) representing the full local habitat range (e.g., wetland and upland) and any

morphological variation that has been noted by workers in the plot, such as variants in bark color or texture or leaf shape or color. These variants may turn out to be cryptic species ((6, 14); Fig. 23). Because forest inventory plots may contain many species represented by single or just a few large trees, additional collections may be needed from outside of the plot in habitats in which the rare plot species may be more abundant.

2. Phenology: Because many tree populations do not reproduce annually, one would ideally make collections over the course of the year and for more than one year in order to obtain representative fruit and floral collections. The collections should be concentrated during seasons in which fruits or flowers are locally most likely.
3. Field observations: Some information about the tagged tree may be obtained from prior inventory data (e.g., DBH, preliminary species assignment, coordinates within plot, tag number, and habitat). The following additional information should be noted for inclusion in the herbarium label: GPS coordinates; date of collection; collector name and collection number; presence of trunk buttresses, bark texture (13). If time permits, noting 5–10 leaf and bark characters for each species can be used to develop basic identification keys for a local flora (15), and can serve to organize photographic resources.
4. Photographic metadata: The field collection presents an opportunity to obtain photographs of fresh flowers and fruits, which contribute valuable information for future species identifications with or without the use of DNA barcodes. In the Gunung Palung flora project in Borneo (16), workers take 10–20 images of each fresh plant, including bark slash, whole twig, twig tip, twig surface, stipules, whole leave (above and below), close up of leaf base underside and petiole, inflorescence, flower (or fruit) at different angles, and longitudinal and transverse sections. Slashing trunks to expose the inner bark is not recommended in forest dynamics plots as it may influence mortality. Each photograph should include a ruler for size scale, and a paper tag with the collector name and collection number (or plot tag number) to avoid confusion about the association of photos and specimens.
5. DNA sampling: From the fresh material collected for the herbarium voucher, select a single young leaf that is neither too tender (the DNA will degrade rapidly in a wilting young leaf) nor excessively damaged by herbivores or covered with epiphylls. Clean a leaf with a dry cloth and cut a 2×2 cm square using scissors. Very little plant tissue (e.g., 20 mg dry tissue) is needed for DNA extraction. A common mistake is to collect too much leaf tissue for DNA sampling using the silica-gel approach. If too much tissue (e.g., an entire leaf) is collected,

the leaf will dry slowly or incompletely, resulting in DNA degradation. Place the tissue sample immediately into a sealed Ziploc sandwich bag or 50-ml Falcon centrifuge tube containing 20 mL of dried silica gel and colored indicator beads. Alternatively, place the sample in a permeable fiber bag (e.g., tea-bag) in a larger box filled with fine silica gel; this prevents fragments of brittle or tender leaves from contaminating the silica gel. Having an excess of silica gel is important for maximum rate of drying. Wipe off scissors with alcohol after each use. Check leaf tissue after one day. If it is brittle and breaks when bent, the silica may be removed and reused if still dry (see color of indicator beads) or baked and reused. Care must be taken not to contaminate samples when reusing silica gel. The dried leaf may be stored indefinitely in a labeled coin envelope inside of an airtight plastic container kept dry with silica gel. There are as yet no standardized protocols for long-term storage of plant tissue for subsequent DNA work. Anecdotal evidence suggests that freezing best preserves DNA in silica-dried leaves.

One alternative to silica gel is to flash freeze the leaves in liquid nitrogen (N). Liquid N is available in many developing countries because it is used to preserve semen for animal breeding. Flash freezing produces more genomic DNA than silica drying and can maintain RNA for transcriptome sequencing. The disadvantage is the difficulty of handling liquid nitrogen tanks in the field, and the expense of long-term storage of frozen material. Liquid N is typically not permitted on flights, so the samples will need to be transported in dry ice, or in a dry shipper. Other alternatives to using silica gel include placing samples in CTAB buffer solution (11), or using Whatman FTA cards (http://www.whatman.com/).

An alternative to using leaf material is obtaining DNA from the vascular cambium (17). Because there may be less need for a plant to invest in defensive secondary compounds in vascular cambium than in leaves, DNA extractions from vascular tissue may be more successful for PCR in some taxa. Pound the gasket-hole puncher into trunk to wood level. Carefully separate the cambium tissue from the inner bark and place in silica gel. Wipe the gasket punch opening with alcohol or bleach to prevent contamination of the next sample.

6. Specimen preservation: The specimen vouchers should be dried on the day of collection when possible, in an arrangement that best demonstrates all of the salient taxonomic characters (e.g., leaf tips, base and underside; stipules, fruits, etc.) (18). Some difficult groups, such as palms, require more specialized arrangement techniques (13). The plant press must be kept tight to prevent wrinkling of material, and retightened through

the course of drying as the material shrinks. The dryer should provide even airflow and temperatures of 35–45°C (12). Rapid drying retains color of specimens but overly high temperatures can produce darkened and brittle specimens.

When collecting in remote areas outside of the field stations, one can layer the fresh collections in newspaper and soak with 90% ethanol (or even methanol, used for lighting lamps, in a pinch). This method will keep the plant parts together until they can be dried, but it produces darkly colored specimens and degrades DNA. For alcohol preserved specimens, fresh leaves should be separately dried with silica gel for use in DNA extraction.

7. Taxonomic sorting: For very large forest inventory plots (e.g., ≥25 ha), sorting all of the designated morphospecies into higher taxonomic ranks can take years, especially if the initial inventory utilized sterile vouchers (19). Key steps in the determination of trees are: (1) collecting "daily vouchers" (either fallen leaves or sterile twigs) for all morphotypes encountered each day, while doing "within-day" matching for trees examined (i.e., "tree 1234 = tree 1245"); (2) matching the daily vouchers to a growing field herbarium collection, assigning field morphotype codes, splitting types where uncertain, and "synonymizing" identical morphotypes with different morphotype codes; (3) determining which taxa can be identified reliably by field crews without further voucher collections (there are always a few common, well-known taxa that "anyone" can spot). This process is time-consuming and tends to slow down because an increasing number of morphotypes have to be checked.

   If the period of sampling is long enough that DNA work can be carried out at the same time, and if enough stems can be sampled for DNA, then sequence data can be used to speed up the matching process (Fig. 2). If the sequence of a new tree can be queried against GENBANK (or otherwise available DNA reference library), or placed in a dynamic, community "guide phylogeny" (automatically rebuilt; see Chapter 19), to find a closely related taxon, then the number of vouchers in the field herbarium to which the new tree's voucher must be compared can be reduced. If there is an exact match of the new tree's sequence to a sequence of a precollected tree, then the first voucher to compare the new tree's voucher with is that of the latter. Because the discriminating power of DNA barcodes in some groups is low (17), we cannot unfortunately expect a direct match of DNA sequence to indicate an exact match of all morphotypes (Fig. 2).

3.2. DNA Barcode Reference Library

The difference between a DNA barcode reference library and a standard DNA sequence database entry (e.g., a standard GenBank

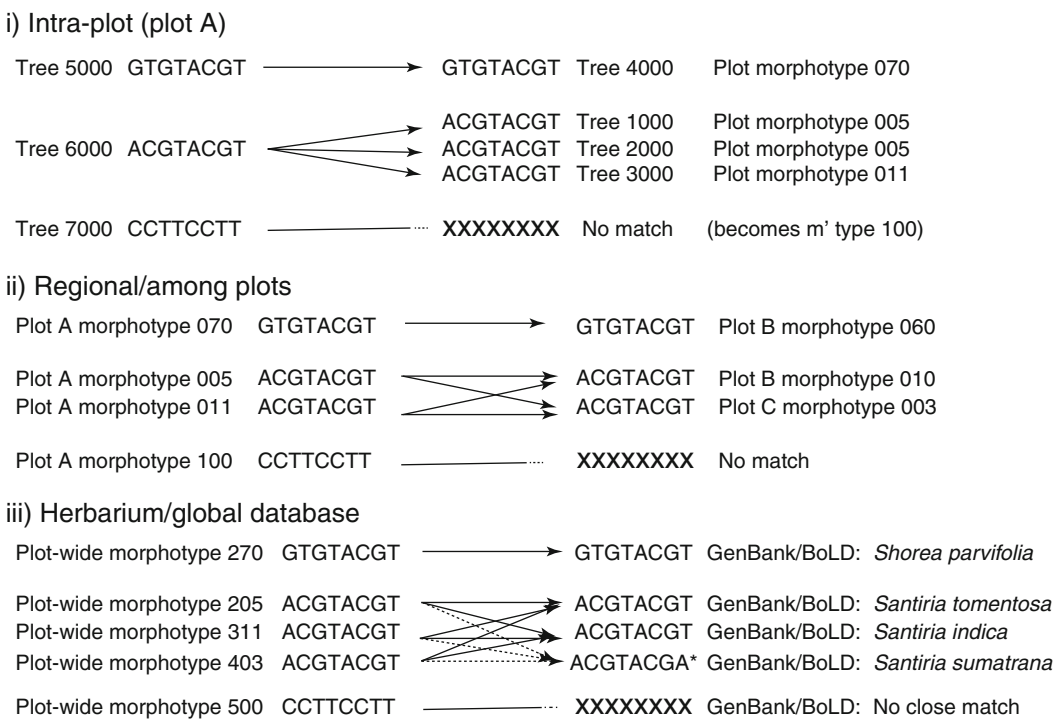

Fig. 2. Hypothetical examples of the use of DNA barcodes for taxonomic management and discovery. (i) Intra-plot matching. DNA from tree 5000 matches only DNA from tree 4000: it is likely that tree 5000 and tree 4000 are the same morphotype and the same species, but a physical comparison is recommended in case two closely related species have identical DNA barcodes. Time saved by using DNA barcodes: only one physical comparison is needed, versus many if no barcodes available. DNA from Tree 6000 matches a DNA sequence from three trees, which has already been found to come from two distinct morphotypes (probably in the same genus): physical comparison is mandatory, to determine the morphotype of Tree 6000. Time saved: only two morphotypes need to be compared with tree 6000. DNA from Tree 7000 does not match DNA from any other tree: it is possible that a physical comparison would find an identical morphotype and reveal a cryptic species, but unlikely. Time saved: physical comparison of tree 6000 is a low priority and could be skipped in some cases. (ii) Inter-plot matching. DNA from plot A morphotype 70 matches only DNA from plot B morphotype 060: it is likely that these morphotypes are the same, and are the same species, but a physical comparison is recommended, in case (a) two closely related but morphologically distinct species have identical DNA barcodes, or (b) there is geographical variation in morphology in one species. In the case of the latter, a taxonomic decision (one species or two) may require herbarium work (see below). Time saved: only one comparison is needed. Identical DNA from plot A distinct morphotypes 005 and 011 matches DNA from plot B morphotype 010 and plot C morphotype 003: thorough physical matching is needed among all four source morphotypes, to determine if there are two, three, or four plot-network-wide morphotypes. Time saved: only these four morphotypes need to be compared, rather than all members of a tentative genus. DNA from plot A morphotype 100 does not match DNA from any other plot morphotype: probably a unique morphotype and species. Time saved: physical comparison of plot A morphotype 100 is a low priority. A final physical review of all morphotypes should be completed, among morphotypes clustered by similar DNA (or by tentative genera, if these have been assigned by field botanists), to determine if there are potentially cryptic species, revealed by different DNA, but having identical morphology. (iii) Herbarium and DNA database matching. DNA from plot-wide morphotype 270 BLASTs to an identical match with *Shorea parvifolia*: it is likely that morphotype 270 is indeed *S. parvifolia*, but with relatively few *Shorea* having ever been sequenced, other *Shorea* may have identical barcodes, hence all *Shorea* in the same section rank should be compared morphologically with vouchers of morphotype 270. If the match is indeed to *S. parvifolia*, then a taxonomic discovery may be made (range expansion, minor morphological variation, etc.). Identical DNA from plot-wide distinct morphotypes 205, 311, and 403 BLASTs to an identical match with *Santiria tomentosa*, *S. indica*, and a close match to *S. sumatrana*. Thorough physical matching in the herbarium (and in monographs) is needed for the three morphotypes, focused on the three possible *Santiria* species, but including all likely *Santiria*, if possible. DNA from plot-wide morphotype 500 does not BLAST to any sequence in any database: the morphotype may be a new species, but more likely it is a species that has been collected before but not sequenced. A herbarium and book search should follow, directed either by a taxonomist's recognition of genus, or starting with taxa with similar DNA sequences.

entry) is that whereas the standard database publishes sequence information at face value, a DNA barcode entry bundles together two hypotheses that must be supported with metadata: (1) that the DNA sequence is accurate, and (2) that the species identification is accurate. The DNA sequence in a DNA barcode reference library must be accompanied by the raw data (chromatogram) so that other researchers can verify that differences in nucleotide sequence between species are robust and not merely sequencing artifacts. The metadata needed to address the taxonomic hypothesis are the herbarium voucher data and accompanying collection information, the most important of which are geographic location, photographs, and collection date. Several data platforms accommodate DNA barcode sequences and metadata. These include the Barcode of Live Database (BOLD) (20) and the DNA barcode entry option of GenBank called "BarSTool". Taxonomic metadata should also be registered in the Global Biodiversity Information Facility (GBIF; http://www.gbif.org), which serves as a repository for biodiversity information, including species ranges.

3.3. Role of the Herbarium

A significant added value of DNA barcode surveys are the associated specimens and genomic DNA that, if properly curated (21), can be used for future generations of biodiversity researchers. It is essential to provide the best quality voucher material (i.e., fertile material) for permanent herbarium curation (most herbaria will not accept sterile or poor-quality specimens). The herbarium provides the infrastructure for exchanging specimens to other institutions so that specialists can make taxonomic determinations and incorporate the specimen information into floras or species descriptions. Herbarium-based curators and systematists can recognize rare or novel taxa, and flag these for additional field collections or observations. Costs for herbarium curation need to be incorporated into research budgets, and collaboration agreements should be established prior to the initiation of a large-scale DNA barcode project. Herbarium staff members are often involved in acquiring research and collection permits, for example, which can be a time consuming and laborious procedure that should be dealt with as early as possible.

3.4. Taxonomic Discovery

The discovery of new species, site records, variants or hybrids involves a comparison of morphological data (morphospecies designation) based on field observations and herbarium vouchers, and the DNA barcode haplotypes (Fig. 2). There are two deviations from the ideal one-to-one relationship between the DNA barcode and the locally defined morphospecies: (1) DNA barcodes are identical across multiple morphospecies, or (2) multiple DNA barcode haplotypes are found within a single putative morphospecies. Since each scenario can arise from different biological causes, these cases require further evaluation (Fig. 2).

Case 1: One DNA barcode for multiple morphospecies. When identical DNA barcodes are found in different morphospecies, it likely reflects a recent speciation history in which mutational differences among species have not yet accrued and sorted. Such is the case in species rich tree genera, such as *Inga* and sections of the genus *Ficus* (22). The genetic discrimination of such taxa will require more variable DNA markers (23); and see discussion). These cases underscore the need to maintain archived DNA for future genotyping.

Shared cpDNA haplotypes may also be explained by hybridization. Hybridization can be detected in several ways including: within a phylogenetic context as incongruence between nuclear and plastid phylogenies, by geographical associations of haplotypes shared across species (24), by levels of genetic admixture with nuclear loci (23), or by morphological intermediates between the putative species in the field. Taxonomic specialists often have a prior idea of the importance of hybridization in their taxa, based on their examination of morphological discontinuities among species.

Case 2: One morphospecies with multiple DNA barcodes. Variant DNA barcodes can be found within species across the geographic range, or even locally in some species (25). This can indicate the existence of morphologically cryptic or semicryptic species, which might have been lumped together as single taxon by field workers (14). In this case, the field workers should revisit the individuals with divergent haplotypes, and carefully examine adult individuals along with nearby seedling and saplings, and collect samples from individuals representing the full range of morphological and ecological variation. If the DNA variation is consistently associated with certain morphological or ecological types, this can provide good evidence of multiple species. These cryptic species can be flagged for further study focused on potential reproductive barriers, such as nonoverlapping phenology and habitat segregation (26). If the two cryptic species are not sister species, then they should also segregate in different nodes within a broadly sampled phylogeny (see Fig. 3).

4. Discussion

Standard cpDNA barcodes will be useful for discriminating species across distant clades and within relatively old clades (e.g., with sister species divergences older than the Pleistocene). We provide the example of *Trema micrantha* species complex (Fig. 3) as an example in which DNA barcodes could be used to discriminate a cryptic species in a long-term forest inventory plot (28).

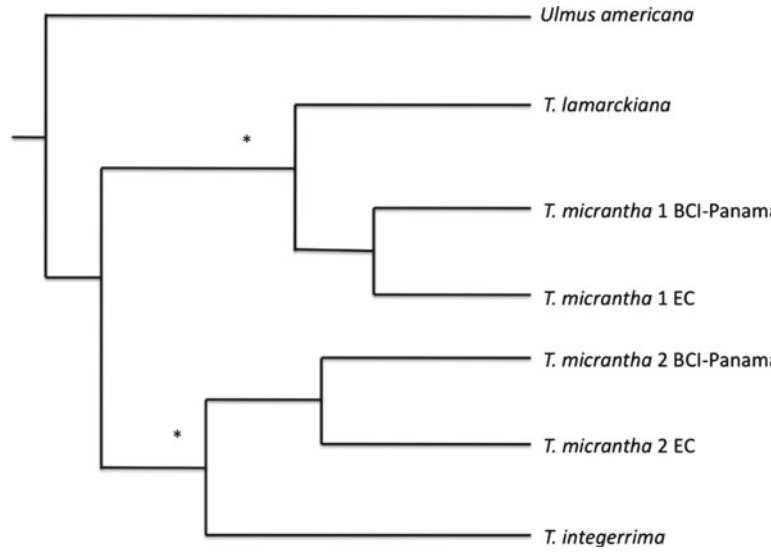

Fig. 3. Example of using DNA barcodes to diagnose cryptic species. In Barro Colorado Island (BCI), Panama, there was thought to be a single species of *Trema*—the common pioneer tree species *Trema micrantha* (27). Molecular studies in the 50 ha plot on BCI revealed highly divergent cpDNA and ITS haplotypes (Dick C, unpublished) among samples, which corresponded with two ecotypes which exhibit ecological differences in light requirement and which can be morphologically distinguished by the color of the endocarp (26). Yesson et al. (2004) showed that showed that *T. micrantha* is a species complex, and the two BCI morphotypes (*T. micrantha* 1 and 2) are not even sister species. Each morphotype is widespread, as indicated by sampling from Ecuador (EC) and form clades with other species with high bootstrap support (*). This phylogeny was adapted from Fig. 2 in Yesson (2004).

The more recently evolved species-rich groups (e.g., *Inga* and *Ficus* sections) may contain enough morphological variation for discrimination in the field, and yet be invariant using the standard plant DNA barcodes. When morphology is not useful for discriminating these species, alternative sets of DNA markers may be used. The nuclear Internal Transcribed Spacer provides more nucleotide substitution variation than most chloroplast DNA and may be amplified using universal primers. Closely related species with recent common ancestors are expected to share many alleles and haplotypes, but their reproductive isolation should be apparent in the form of distinct allele frequencies among syntopic (co-occurring in the same habitat) populations of the putative species. This requires a population genetics approach.

The forest inventory plots provide an excellent system in which to detect reproductive barriers based on genetic differentiation using tools of population genetics because (1) populations of the target species are already mapped and available for analyses and (2) because the species occur in the same locale, the genetic differentiation analysis will not be confounded by differentiation due to isolation by distance processes.

Microsatellite DNA markers (also known as simple sequence repeats or SSRs) are the most commonly used DNA markers for such analyses because of the high rate of mutation and allele richness within populations. Microsatellites are typically isolated from anonymous regions of the nuclear genome. However, because the primers that are designed from the flanking nucleotide sequences are also variable, the microsatellite markers are often species-specific or transferable only to very closely related species. It is not feasible to develop novel microsatellite DNA markers for every potential cryptic species pair. When working within families, an alternative method is to develop microsatellite DNA markers from Expressed Sequence Tags (ESTs). ESTs are short DNA fragments of expressed genes obtained from messenger RNA (mRNA). Although the mRNAs code for proteins, they contain untranslated regions (UTRs) at the 3′ and 5′ ends with SSRs at a frequency of 1–2% (29). Because EST-SSR loci are adjacent to coding sequences, highly conserved PCR primers can be designed, which are transferable across species, genera, and even higher-level taxa (29). EST-SSRs have an additional advantage over anonymous nSSRs in that they generally do not produce null alleles (unamplified alleles) because of their highly conserved priming sites. EST-SSRs can be mined from online EST databases using Web-based bioinformatics search engines. There are currently more than 52 million ESTs in GenBank, including thousands from important and species rich tropical tree families, such as Fabaceae, Rubiaceae, and Lauraceae.

The multilocus dataset for multiple species can be analyzed using Bayesian clustering approaches that estimate the most likely number of genetic demes (K) in the sample (this can be done using the program STRUCTURE) (23, 30). If, for example, five morphospecies represent five distinct species, the analysis should infer $K=5$ demes and assign all individuals to their morphospecies-defined deme. The existence of demes in forest plots is indicative of reproductive isolation (i.e., true species under a biological species concept) because there is not sufficient distance to impede gene flow due to geographic distance. The sample size often used for population genetic analyses is approximately 30 individual per species (to obtain allele frequencies), using ca. 10 SSR loci (for multiple independent estimates of demes). Individuals should be sampled at spaced intervals (e.g., 50 m) throughout the plot to avoid sampling of close relatives.

ESTs can also be a source of phylogenetically informative introns. Although introns are spliced from the mRNA, the EST can be compared to known genomes (e.g., *Arabidopsis thaliana* or *Populus trichocarpa*) to determine which ESTs span introns. From these, Exon Primed Intron Crossing (EPIC) markers can be developed (31). EPIC markers are expected to amplify nuclear introns broadly across higher-level taxa because of their highly conserved priming regions. Markers such as these will be useful for distinguishing among closely related species, and for developing phylogenies for establishing species relationships.

In summary, we see great potential for DNA analyses to assist in the management of taxonomic entities in species-rich forest inventory plots, and in the discovery of new species. We can imagine a time when a multilocus library of DNA sequences existed for all named species, with an estimate of sequence variation within each species, and when we could affordably sequence millions of base pairs for each individual in the plot using pyrosequencing. We could then match trees to local plot taxa, and to named species, without ever consulting physical vouchers (not that such a DNA-only approach would necessarily be desirable). Significantly original sequences would then almost certainly indicate species new to science. However, we are of course far from having these data available, and so DNA barcodes must be considered an additional valuable source of data in our taxonomic work, to be used in dialog with physical vouchers, rather than a goal in themselves.

Acknowledgments

Some of the methods were derived from research supported by the National Science Foundation (DEB awards 0640379 to CD, and 1020868 to CW) and the Center for Tropical Forest Sciences. We thank John Kress and David Erickson for the invitation and for useful ideas for the paper.

References

1. Fine PVA, Ree RH (2006) Evidence for a time-integrated species-area effect on the latitudinal gradient in tree diversity. Am Nat 168: 796–804
2. Jørgensen PM, León-Yánez S (1999) Catalogue of the vascular plants of Ecuador. Missouri Botanical Garden, St. Louis, MO
3. Balslev H, Valencia R, Paz y Miño G, Christensen H, Nielsen I (1998) Species count of vascular plants in one hectare of humid lowland forest in Amazonian Ecuador. Forest biodiversity in North, Central and South America, and the Caribbean, research and monitoring. In: Dallmeier F, Comiskey JA (eds) Man and the biosphere series, vol 21. UNESCO, Paris, pp 585–594
4. Ruokolainen K, Tuomisto H, Kalliola R (2005) Landscape heterogeneity and species diversity in Amazonia. In: Bermingham E, Dick CW, Moritz C (eds) Tropical Rainforests, Past, Present and Future. University of Chicago Press, Chicago, pp 251–270
5. Prance GT, Beentje H, Dransfield J, Johns R (2000) The tropical flora remains undercollected. Ann Mo Bot Gard 87:67–71
6. Dick CW, Kress WJ (2009) Dissecting tropical plant diversity with forest plots and a molecular toolkit. Bioscience 59:745–755
7. Bebber DP, Carine MA, Wood JRI et al (2010) Herbaria are a major frontier for species discovery. Proc Natl Acad Sci U S A. doi:10.1073/pnas.1011841108
8. Mori SA, Prance GT (1987) A guide to collecting lecythidaceae. Ann Mo Bot Gard 74: 321–330
9. RAINFOR (Amazon Forest Inventory Network) http://www.geog.leeds.ac.uk/projects/rainfor/pages/project_eng.html. Last accessed on 28 Feb 2011
10. Laman TG (1995) Safety recommendations for climbing rain-forest trees with single rope technique. Biotropica 27:406–409
11. Colpaert N et al (2005) Sampling tissue for DNA analysis of trees: trunk cambium as an alternative to canopy leaves. Silvae Genetica 54:265–269
12. Blanco MA et al (2006) A simple and safe method for rapid drying of plant specimens using forced-air space heaters. Selbyana 27:83–87

13. Leisner R (Field Techniques used by the Missouri Botanical Garden) http://www.mobot.org/mobot/molib/fieldtechbook/welcome.shtml. Last accessed on 28 Feb 2012 (Missouri Botanical Garden, St. Louis, MO)
14. Janzen DH et al (2009) Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. Mol Ecol Resour 9:1–26
15. Kress WJ (2004) Plant floras: how long will they last? A review of flowering plants of the Neotropics. Am J Bot 91:2124–2127
16. Webb CO, Slik JWF, Triono T (2010) Biodiversity inventory and informatics in Southeast Asia. Biodivers Conserv 19:955–972
17. Gonzalez M-A, Baraloto C, Engel J et al (2009) Identification of Amazonian trees with DNA barcodes. PLoS Biol 4:e7483
18. Herbarium University of Florida (Preparation of plant specimens for deposit as herbarium vouchers) http://www.flmnh.ufl.edu/herbarium/voucher.htm#Identification. Last accessed on 28 Feb 2012
19. Condit R (1998) Tropical forest census plots: methods and results from Barro Colorado Island. Panama and a comparison with other plots, Springer-Verlag, Berlin
20. Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). Mol Ecol Notes 7:355–364
21. Savolainen V, Reeves G (2004) A plea for DNA banking. Science 304:1445
22. Kress WJ 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
23. Duminil J, Caron H, Scotti I, Cazal SO, Petit RJ (2006) Blind population genetics survey of tropical rainforest trees. Mol Ecol 15:3505–3513
24. Saeki I, Dick CW, Barnes BV, Murakami N (2011) Comparative phylogeography of red maple (*Acer rubrum* L.) and silver maple (*A. saccharinum* L.): impacts of habitat specialization, hybridization and glacial history. J Biogeogr 38:992–1005
25. Dick CW, Heuertz M (2008) The complex biogeographic history of a widespread tropical tree species. Evolution 62:2760–2774
26. Silvera K, Skillman JB, Dalling JW (2003) Seed germination, seedling growth and habitat partitioning in two morphotypes of the tropical pioneer tree *Trema micrantha* in a seasonal forest in Panama. J Tropical Ecol 19:27–34
27. Croat TB (1978) Flora of Barro Colorado Island. Stanford University Press, Stanford, CA
28. Yesson C, Russell SJ, Parrish T et al (2004) Phylogenetic framework for Trema (Celtidaceae). Plant Syst Evol 248:85–109
29. Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. Heredity 99:125–132
30. Pritchard JK, Stephens JC, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
31. Li C, Riethoven JM, Ma L (2010) Exon-primed intron-crossing (EPIC) markers for non-teleost fishes. BMC Evol Biol 10:90