BIOTROPICA
THE JOURNAL OF TROPICAL BIOLOGY AND CONSERVATION

BIOTROPICA 40(4): 405–408 2008 10.1111/j.1744-7429.2008.00426.x

DNA Barcoding—a Windfall for Tropical Biology?

W. John Kress[1] and **David L. Erickson**

Department of Botany, MRC-166, National Museum of Natural History, Smithsonian Institution, P.O. Box 37012, Washington, DC 20013-7012, U.S.A.

'I am standing in a Costa Rican Rain forest', writes tropical ecologist and conservationist Dan Janzen (2005). 'There are a thousand species of plants within a long stone's throw. Nearly every one of them is a described species with a proper scientific name, a handle that you can plug into Google and come up with something. . . . And I cannot identify a single species. Imagine what it would do to any and all aspects of human interactions with wild plants if you could walk up to any plant anywhere—seedling, sapling, 40 m tree, grass, root, pressed leaf, or fallen log—and know in a few seconds its scientific name'.

Janzen, one of tropical biology's more ardent advocates, poetically describes the taxonomic impediment that exists today for many ecologists and evolutionary biologists working in the field: determining the correct species identification for any plant sample in a fast and repeatable fashion. In fact Janzen knew when he wrote this passage that a rapid and accurate method is now being developed and refined for the quick identification of plant species based on extracting DNA sequence from a tiny tissue sample from any part of a plant. Appropriately called 'DNA barcoding', referring to the coded labels one finds on commercial products, DNA barcodes consist of a standardized short sequence of DNA between 400 and 800 base pairs long that can be easily generated and characterized for all species on the planet (Savolainen *et al.* 2005). These genetic barcodes will be stored in an open access digital library that can be used to compare the DNA barcode sequence of an unidentified sample from the field, garden, or market by matching it to a known sequence with an associated species name in the data base. DNA barcoding has the potential to greatly advance our access to the collective knowledge of biodiversity and in turn our understanding of Nature. By harnessing advances in molecular genetics, sequencing technology, and bioinformatics, DNA barcoding will allow users to quickly and cheaply recognize known species and retrieve information about them. It may also speed the discovery of the thousands of species yet to be named. Barcoding, if developed sufficiently, will be a vital new tool for appreciating and managing the Earth's immense and changing biodiversity (Cowan *et al.* 2006). And most of that diversity is in the tropics.

Some of the basic questions about tropical ecosystems that biologists have been attempting to answer over the last century require an estimate of species diversity. Why do tropical biomes have greater species diversity than temperate zone habitats? How is this species richness maintained in time and space? What are the ecological and evolutionary processes that influence species composition in tropical forests? Answers to some of these broad questions initially require accurate assessments of the level of species diversity in a particular habitat. These assessments must have accurate species identifications and, ideally, a uniform species definition. As tropical biologists, we all know how difficult it is to obtain accurate identification of organisms in species-rich tropical forests. DNA barcoding will provide a new tool that will supplement traditional methods and allow us to more independently assess this overwhelming species diversity.

WHAT IS DNA BARCODING?

A DNA barcode, in its simplest definition, is one or more short gene sequences taken from a standardized portion of the genome used to identify species. The use of such short DNA sequences for biological identifications was first proposed by Paul Herbert and colleagues (2003a, 2004a) with the ultimate goal of quick and reliable species-level identifications across all forms of life, including animals, plants, and microorganisms. Up to the present, the concept of a universally recoverable segment of DNA that can be applied as an identification marker across species has been most successfully applied to animals (Hebert *et al.* 2004b). For plants, DNA-based identifications, although not called 'barcodes', have been used to describe extinct herbivore diets (Poinar *et al.* 1998, Hofreiter *et al.* 2000), to identify species of wood (Deguilloux *et al.* 2002) and species used in herbal supplements (Zerega *et al.* 2002), and to associate roots in Texas caves with trees growing on the surface (Jackson *et al.* 1999). Despite such success in using DNA sequence data for plant identification a standard DNA barcode for plants has not yet been accepted by the botanical community (see below). However, several recent broad screenings of gene regions in the plastid genome, some of which may have great potential for plant species identification (Chase *et al.* 2007, Kress & Erickson 2007, Lahaye *et al.* 2008), suggest that a standard barcode for plants may soon be accepted.

At the minimum, three criteria must be addressed in identifying a gene region as appropriate to be a DNA barcode: (1) significant species-level genetic variability and divergence; (2) short sequence length to facilitate DNA extraction and amplification; and (3) universal PCR primers. For most groups of animals, a portion of

Received 12 December 2007; revisions accepted 29 March 2008.
[1]Corresponding author; e-mail: kressj@si.edu

the mitochondrial gene for cytochrome c oxidase 1 (COI; Hebert *et al.* 2003a, b, 2004a) has been identified as a species-level barcode (see http://www.barcoding.si.edu/). An important rationale for using short sequences also resides in the need to obtain useful data from potentially degraded samples found in museum specimens. COI has been shown to fit the three criteria in the great majority of animal taxa to which it has been applied (Hebert *et al.* 2003b, but see Meyer & Paulay 2005). Unfortunately, the selection of a universally applied DNA barcode for plants has not been as straightforward as was the case for animals.

The primary reason that DNA barcoding was not immediately applied to plants is that plant mitochondrial genes, due to their low rate of sequence change, are poor candidates for species-level discrimination. The divergence of COI coding regions among families of flowering plants has been documented to be extremely limited (only a few base pairs across 1.4 kb of sequence; Cho *et al.* 2004). Because COI (or any other mitochondrial gene region) does not work as a barcode in plants, botanists have been testing other genetic markers. In the last 2 yr, a number of candidate gene regions have been suggested as a barcode for plants, most of them located in the chloroplast genome (coding regions *accD*, *matK*, *ndhJ*, *rpoB2*, *rpoC1*, and *ycf5*, Chase *et al.* 2007, Lahaye *et al.* 2008; coding region *rbcL*, Kress & Erickson 2007; *trnL* intron, Taberlet *et al.* 2006; and noncoding spacer *trnH-psbA*, Kress *et al.* 2005, Kress & Erickson 2007). The Internal Transcribed Spacer, or ITS, located in the nuclear genome of plants has also shown partial promise as a plant DNA barcode (Kress *et al*. 2005, Sass *et al.* 2007). One of the most promising candidates for an effective plant DNA barcode appears to be a combination of two or more gene regions (Rubinoff *et al.* 2006, Chase *et al*. 2007) that incorporates several coding loci or a noncoding spacer and a coding gene. Such combinations include the three coding genes *matK*, *rpoB*, and *rpoC1* (Chase *et al.* 2007) and the spacer region *trnH-psbA* plus coding gene *rbcL* (Kress & Erickson 2007). The controversy over selecting a plant barcode region stems from basic disagreements on such issues as the purpose of barcoding, the degree of universality of barcoding, and bioinformatics issues, including ease of sequence alignment and retrieval (Kress & Erickson 2008).

From its inception, the primary use of DNA barcodes has been for identification. As a research tool for taxonomists, barcoding will aid in identification by expanding species diagnoses to cover all life-history stages of an organism, dimorphic sexes, damaged specimens, gut contents, scats, and fecal samples. In addition, systematists will be able to quantify the consistency of their species definitions with a measure of genetic variability based on the sequence data. For the applied users of taxonomy, barcoding will be a tool to identify regulated species, including invasives, as well as to test the identity and purity of biological products, such as herbal medicines and supplements. As a biodiversity discovery tool, barcoding will help to flag species that are potentially new to science, especially undescribed and cryptic species (see Hebert *et al.* 2004b). For tropical biologists, who work in forests in which many of the species are still unknown, this new discovery tool will be exceptionally useful. If a plant barcode is chosen that has both depth and breadth of sequence variability, the universal region could assist in constructing community phylogenies (Kembel & Hubbell 2006, Kress & Erickson 2007). Of course today the process of sequencing DNA is still restricted to laboratories. However, as discussed at the second *International Barcode of Life Conference* held in Taipei in September of 2007 it is anticipated that sequencing will eventually be more accessible and even portable.

The enthusiasm for the potential uses of barcoding within both the scientific and applied community has resulted in the formation of the *Consortium for the Barcode of Life* (CBOL; see http://www.barcoding.si.edu/), which is housed at the Smithsonian's National Museum of Natural History in Washington, DC., and funded by the Alfred P. Sloan Foundation. CBOL has succeeded in building a community of museums, botanic gardens, aquaria, and research institutions that now numbers over 150 institutes from 45 countries across six continents. The Consortium is charged with coordinating barcoding activities around the world and promoting a data base of documented and vouchered reference sequences to serve as a universal DNA barcode library for all life.

WHY THE CONTROVERSY?

DNA barcoding follows the same principle as the basic taxonomic practice of identification by associating a name with a specific reference collection in conjunction with a functional understanding of species concepts. As with the introduction of any new method of analysis in science, some controversy and concern has arisen about the feasibility and utility of DNA barcoding in taxonomy (Holmes 2004). A number of taxonomists appear to be skeptical of new methodologies that may further the 'Linnaean enterprise' (*i.e.*, inventory, identification, and classification of life), but threaten the field of taxonomy. Some are particularly concerned that new technologies might be substituted for the taxonomic specialists doing their job working directly with specimens. Others believe that these new techniques will be misused and give faulty results.

These misconceptions arise for a number of reasons, such as (1) confusing DNA barcoding with DNA taxonomy, which defines species strictly by their level of genetic divergence; (2) confounding 'service identifications' through DNA barcoding with the entire field of taxonomy when it is only one aspect of it; and (3) believing that any new tool, such as DNA barcoding, will replace the need for taxonomic specialists or at least siphon off all of their funds and professional positions (Kress 2004). The last concern so far has not proven to be true. New funding has been found to promote and advance barcoding that would not have been allocated to taxonomy if not for this new technology. The easier it is for end-users to employ good taxonomic data for identification, the more the field of taxonomy and taxonomists will be appreciated for their skills and knowledge. The practical application of DNA barcoding will require the assembly of a data base where sequences are linked to established species. Taxonomists will form the vanguard in the development of this data base because they are essential to assigning species designations to vouchers that are sequenced and submitted to the data base. In addition, they will provide nonsequence taxonomic information about the species that will be added to the

data base either directly or through such portals as the *Encyclopedia of Life* (http://www.eol.si.edu). The need for taxonomists to build the barcode data base may very well give rise to a new generation of individuals who will be trained in part to help identify and collect the specimens upon which the DNA barcode library will be built. With regard to the public, interest in nature, biodiversity, and the environment by nonprofessionals is soaring and the demand for effective field guides and easy identification aids is at an all time high (Gorman 2004; Janzen 2004, 2005). If more taxonomic information is available to nonspecialists to use for species identifications, more appreciation and respect will be accorded to the taxonomists who supplied that information. And in the long run, respect for Nature will proportionally increase.

In some cases, DNA barcoding has been shown to just not work (*e.g.*, Hickerson *et al.* 2006, Meier *et al.* 2006, Elias *et al.* 2007). In addition, some concern is warranted that barcodes will give poor results or faulty identifications because of the complications of ancestral polymorphisms, hybridization, and/or introgression (Moritz & Cicero 2004). So far, these possible pitfalls have not slowed the enthusiasm for the potential application of DNA barcoding to both plants and animals.

EARLY APPLICATIONS OF PLANT DNA BARCODING

As stated earlier, it is generally agreed that a plant barcode will require a combination of several gene regions to adequately provide the required level of discrimination. With this in mind, a number of projects have been initiated as a proof of concept for plant barcoding based on a two-locus model. One project, to build a barcode library of over 700 species of the world's most important medicinal plants (W. J. Kress & E. Erickson, unpubl.), is nearly complete and will allow us to test the identity and purity of plant-based medicines and herbals, such as ginseng, ginkgo, echinacea, and St. John's wort, sold in commercial markets to local consumers. In another case, the potential to use DNA barcodes in the regulation of the trade in endangered tropical plant species, especially orchids (Lahaye *et al.* 2008), has been demonstrated. These efforts illustrate the suite of uses of DNA barcodes with substantial economic and social value.

A larger project, which makes DNA barcoding more directly relevant to tropical biology, has just begun. Sponsored by the Smithsonian Institution, a complete barcode survey of the woody plant taxa found in the 50-ha 'Forest Dynamics Plot' on Barro Colorado Island in Panama is now underway (W. J. Kress, E. Erickson, E. Bermingham, and O. Sanjur, unpubl.). This genetic inventory of the plant species will provide an exciting complement to the diversity of studies that have arisen from the mapping and periodic demographic censuses of the plot, which have shown that tropical forest populations are extremely dynamic and responsive to climate change. To date, over 95 percent of the species in the plot have shown a unique genetic signature with the *rbcL*/*trnH-psbA* plant barcode combination. This two-locus DNA marker can now be used to accurately identify herbivore host plants lacking diagnostic flowering features, disarticulated fruits, seeds and seedlings, as well as plant fragments in animal guts and scats. It has also been possible to use the sequence data from this two-locus barcode to develop a measure of the phylogenetic structure of the woody plant community in the plot, thereby allowing an assessment of the evolutionary component of forest plant traits that influence ecosystem function (Kembel & Hubbell 2006; J. Wright, pers. comm.). The Barro Colorado Island 50-ha plot is only one of 20 sites located in 15 tropical countries around the world that are being monitored by the *Center for Tropical Forest Science* (http://www.ctfs.si.edu/doc/index.php). A new initiative at the Smithsonian, called the *Smithsonian Institution Global Earth Observatories* (SIGEO), will monitor the 3.5 million trees at these 20 sites as well as adding a number of sites located in the temperate zone to quantitatively determine how trees and forest ecosystems are responding to the Earth's changing climate. A complete plant barcode census is planned for all the woody plants in these sites, which together account for over 12 percent of all known tropical tree species. This effort will require significant human and financial investment to complete, but will provide not only accurate identifications of the species in these plots, which can be applied to multiple investigations, but will also allow new genetic comparisons to be made among species and communities throughout the various ecosystems across the tropics. A similar effort is now being initiated to barcode the two thousand species of vascular plants at La Selva Biological Station in Costa Rica run by the Organization for Tropical Studies (D. Stone & E. Losos, pers. comm.). As has already been demonstrated with animal barcodes (Hebert *et al.* 2004b) a spectrum of new plant species will undoubtedly be discovered as a result of these broad genetic inventories in the tropics.

In summary, DNA barcoding holds promise for numerous applications and biological disciplines in the tropics. Today, DNA barcodes are necessarily locus-based and relatively short in length. As sequencing technology advances it is easy to foresee that the DNA barcodes of the future may be based on not just single or multiple genes, but on entire genomes, which will allow completely new biological questions to be posed and answered (van Straalen & Roelofs 2006). These genetic signatures of species, which at the minimum ensure accurate identifications, have the potential to provide much more information about a plant, including insights into its geographic origin, placement in the Tree of Life, maternal and paternal lineages, as well as its genomic structure (Kress & Erickson 2008). Still, many hurdles need to be overcome before barcoding will be practical for tropical scientists, let alone the public. But if the development of geographic positioning systems is a model for progress on a handheld DNA sequencer, one can imagine that a sequencing machine that now fits on a table top may eventually slip into a backpack if not a pocket. Regardless of the future scale of the technology, it is important to start populating the DNA barcode library now. Indeed, in the not too distant future Dan Janzen's vision of a method for instant plant identification may seem quaint.

LITERATURE CITED

Chase, M. W., R. S. Cowan, P. M. Hollingsworth, C. Van Den Berg, S. Madriñán, G. Petersen, O. Seberg, T. Jørgsensen, K. M. Cameron, M. Carine, N. Pedersen, T. A. J. Hedderson, F. Conrad, G. A.

Salazar, J. E. Richardson, M. L. Hollingsworth, T. G. Barraclough, L. Kelly, and M. Wilkinson. 2007. A proposal for a standardised protocol to barcode all land plants. Taxon 56: 295–299.

Cho, Y., J. P. Mower, Y.-L. Qiu, and J. D. Palmer. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proc. Natl. Acad. Sci. USA 101: 17741–17746.

Cowan, R. S., M. W. Chase, W. J. Kress, and V. Savolainen. 2006. 300,000 species to identify: Problems, progress, and prospects in DNA barcoding of land plants. Taxon 55: 611–616.

Deguilloux, M.-F., M.-H. Pemonge, and R. J. Petit. 2002. Novel perspectives in wood certification and forensics: Dry wood as a source of DNA. Proc. R. Soc. Lond. B 269: 1039–1046.

Elias, M., R. I. Hill, K. R. Willmott, J. Mallet, and C. D. Jiggins. 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proc. R. Soc. Lond. B 274: 2881–2889.

Gorman, J. 2004. So many field guides you need a guide. The New York Times, July 20, p. D3.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. deWaard. 2003a. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B 270: 313–321.

Hebert, P. D. N., S. Ratnasingham, and J. R. deWaard. 2003b. Barcoding animal life: Cytochrome *c* oxidase subunit 1 divergences among closely related species. Proc. R. Soc. Lond. B. 270 (suppl.): S96–S99.

Hebert, P. D. N., M. Y. Stoeckle, T. S. Zemlak, and C. M. Francis. 2004a. Identification of birds through DNA barcodes. PLoS Biol. 2: e312.

Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs. 2004b. Ten species in one: DNA barcoding reveals cryptic species in the Neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. USA 101: 14812–14817.

Hickerson, M. J., C. P. Meyer, and C. Moritz. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. Syst. Biol. 55: 729–739.

Hofreiter, M., H. N. Poinar, W. G. Spaulding, K. Bauer, P. S. Martin, G. Possnert, and S. Pääbo. 2000. A molecular analysis of ground sloth diet through the last glaciation. Mol. Ecol. 9: 1975–1984.

Holmes, R. 2004. Barcode me. New Scientist 26 June, 32–35.

Jackson, R. B., L. A. Moore, W. A. Hoffmann, W. T. Pockman, and C. R. Linder. 1999. Ecosystem rooting depth determined with caves and DNA. Proc. Natl. Acad. Sci. USA 96: 11387–11392.

Janzen, D. H. 2004. Now is the time. Philos. Trans. R. Soc. London B 359: 731–732.

Janzen, D. H. 2005. Foreword: How to conserve wild plants? Give the world the power to read them. *In* G. A. Krupnick, and W. J. Kress (Eds.). Plant conservation: A natural history approach, pp. ix–xiii. University of Chicago Press, Chicago, Illinois.

Kembel, S. W., and S. P. Hubbell. 2006. The phylogenetic structure of a Neotropical forest tree community. Ecology 87(suppl.): S86–S99.

Kress, W. J. 2004. Paper floras: How long will they last? A review of *Flowering Plants of the Neotropics*. Am. J. Bot. 91: 2124–2127.

Kress, W. J., K. J. Wurdack, E. A. Zimmer, L. A. Weigt, and D. H. Janzen. 2005. Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. USA 102: 8369–8374.

Kress, W. J., and D. L. Erickson. 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. doi:10.1371/journal.pone.0000508.

Kress, W. J., and D. L. Erickson. 2008. DNA barcoding: Genes, genomics, and bioinformatics. Proc. Natl. Acad. Sci. USA 105: 2761–2762.

Lahaye, R., M. Van Der Bank, D. Bogarin, J. Warner, F. Pupulin, G. Gigot, O. Maurin, S. Duthoit, T. G. Barraclough, and V. Savolainen. 2008. DNA barcoding the floras of biodiversity hotspots. Proc. Natl. Acad. Sci. USA 105: www.pnas.org_cgi_doi_10.1073_pnas.0709936105

Meier, R., K. Shiyang, G. Vaidya, and P. K. L. Ng. 2006. DNA barcoding and taxonomy in diptera: A tale of high intraspecific variability and low identification success. Syst. Biol. 55: 715–728.

Meyer, C. P., and G. Paulay. 2005. DNA Barcoding: Error rates based on comprehensive sampling. PLoS Biol 3: e422.

Moritz, C., and C. Cicero. 2004. DNA barcoding: Promises and pitfalls. PLoS Biol 2: 1529–1531.

Poinar, H. N., M. Hofreiter, W. G. Spaulding, P. S. Martin, B. A. Stankiewicz, H. Bland, R. P. Evershed, G. Possnert, and S. Pääbo. 1998. Molecular coproscopy: Dung and diet of the extinct ground sloth *Nothrotheriops shastensis*. Science 281: 402–406.

Rubinoff, D, S. Cameron, and K. Will. 2006. Are plant DNA barcodes a search for the Holy Grail? Trends Ecol. Evol. 21: 1–2.

Savolainen, V., R. S. Cowan, A. P. Vogler, G. K. Roderick, and R. Lane. 2005. Towards writing the encyclopedia of life: An introduction to DNA barcoding. Phil. Trans. Roy. Soc. London Ser. B Biol. Sci. 360: 1850–1811.

Sass, C., D. P. Little, D. Wm. Stevenson, and C. D. Specht. 2007. DNA barcoding in the Cycadales: Testing the potential of proposed barcoding markers for species identification of cycads. PLoS ONE 2: e1154. doi: 10.1371/journal.pone.0001154

Taberlet P., E. Coissac, F. Pompanon, L. Gielly, C. Miquel, A. Valentini, T. Vermat, G. Corthier, C. Brochmann, and E. Willerslev. 2006. Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding. Nucleic Acids Res e1–8 doi:10.1093/nar/gkl938.

van Straalen N-M, and D. Roelofs. 2006. An introduction to ecological genomics. Oxford University Press, London, UK.

Zerega, N. J. C., S. Mori, C. Lindqvist, Q. Zheng, and T. J. Motley. 2002. Using amplified fragment length polymorphisms (AFLP) to identify black cohosh (*Actaea racemosa*). Econ. Bot. 56: 154–164.