Update TRENDS in Ecology and Evolution Vol.21 No.1 January 2006

Research Focus

Are plant DNA barcodes a search for the Holy Grail?

Daniel Rubinoff[1], Stephen Cameron[2] and Kipling Will[3]

[1]Department of Plant and Environmental Protection Sciences, 310 Gilmore Hall, 3050 Maile Way, University of Hawaii, Honolulu, HI 96822, USA
[2]Department of Integrative Biology, Brigham Young University, Provo, UT 84602, USA
[3]Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA

In a recent study, Kress *et al.* compared two plant genomes to seek out plant DNA barcodes. Two promising markers balanced the variability that is needed to distinguish species with conserved primer regions that enable universal amplification. Although this study is the most rigorous effort to date, problems from earlier barcoding efforts, such as the use of non-evolutionary species concepts and differential sorting of genes and species, could reemerge. Single-gene barcoding might not be universally effective owing to inherent inaccuracies. Kress *et al.* suggest the use of multiple genes, reflecting an integrated approach that is likely to be the best answer to identifying species quickly and accurately.

What is DNA barcoding?

The diversity of life, as measured by numbers of species, is staggering. Taxonomists could take decades to describe the estimated 10 million–15 million species [1] using current methods of description and publication. DNA barcodes have been proposed as a shortcut that would provide species identifications and as a way to accelerate the discovery of new species. DNA barcodes are short segments (~800 bp) of a gene sequence that evolve fast enough to differentiate species, but have flanking regions that are sufficiently conserved to enable the barcode region to be serviced by universal primers. Barcodes can identify previously described species, but to impact the fundamental crisis facing biodiversity, they will need to provide a means to address the major issue of species that are undescribed and completely unknown. Until recently, plants had been largely left out of barcode efforts because the gene of choice for animals, *COI*, as with all mitochondrial genes, does not evolve at an appropriate rate for species-level discernment in plants. In a new paper, Kress *et al.* [2] attempt to remedy this inequity with an analysis and comparison of several genetic markers that they suggest could serve as possible substrates for a plant kingdom barcode effort.

Barcoding plants

Kress *et al.* [2] present the most logical search for a barcoding marker offered to date. They compared the utility of two plastid genomes for two species (tobacco *Nicotiana tobacum* and deadly nightshade *Atropa belladonna*) and then tested candidate barcode regions on 99 species in 80 genera from 53 plant families. Their search suggested two promising regions, a plastid nuclear intergeneric spacer (*trnH-psbA*) and the internal transcribed spacer for RNA-coding nuclear genes (*ITS*). The *trnH-psbA* region exhibited high divergence levels, but is only 450 bp long, shorter than typical animal barcodes that are 600–800 bp. Although *trnH-psbA* was the most variable plastid region in angiosperms and easily amplified across the group, its short length might not provide enough data for universal use. The second candidate, *ITS*, was found to evolve faster than many plastid regions, including the *rbcL* chloroplast. Kress *et al.* [2] mention that, although *rbcL* was considered previously as a source for barcoding and is the plastid locus most commonly sequenced for plant systematic studies, it evolves too slowly to be of broad use for species level identifications. Additionally, *ITS* appears to have largely, although not universally, conserved primer regions, which are essential to broad-based species-level identification. However, *ITS* failed to amplify for 12% of herbarium samples, and was of poor quality for many others. Amplification of DNA from herbarium specimens is important for barcoding because it is necessary to confirm many identifications, particularly of rare or taxa that are presumed extinct. The combination of *ITS* and *trnH-psbA* might be the best candidate for a plant barcode, but neither is sufficient on its own to identify or define species.

Taxonomy, phylogenetics and species identifications

Advocates say barcoding is a 'practical, standardized, species-level identification tool for biodiversity assessment, life history and ecological studies, and forensic analysis' [2], although there have been criticisms of the distance-based phylogenetic approach used to identify barcode samples [3,4]. Some barcode advocates have sought to avoid phylogenetically based criticisms of their methodology by claiming that barcoding is for identifications and not phylogenies [1,2]; however, barcoding currently requires the construction of relationships between sequences that are attributed to identified species and sequences from unknown samples using phylogenetic methods [5,6]. This inconsistency is evident when Kress *et al.* discuss the phylogenetic utility of the genes they are evaluating: 'Species-level discrimination and technical ease have been validated in most phylogenetic studies that employ *ITS*…' yet maintain that barcoding has nothing to do with phylogenetics. In so doing, they avoid addressing, and benefiting from, the body of phylogenetic literature that suggests that a single, small gene sequence is generally inadequate for phylogenetic analysis and

Corresponding author: Rubinoff, D. (rubinoff@hawaii.edu).
Available online 14 November 2005

might fail for species identification [7]. The use of non-evolutionary, phenetic distance methods as a means to identify unknown samples confuses identification with definition. It is unrealistic to think that *COI*, or any single character system, will be sufficiently accurate to define all species. Therefore, to have meaningful species, they should be studied in the evolutionary context of related species first. Only after species are defined can a shortcut to identification be used; for example, a differential key or perhaps even a single 'barcode' sequence. However, such a procedure does not conform to the selling point for barcodes as not only a shortcut method for well-known groups, but also a novel way to identify the many undescribed species.

Over the past 50 years, it has become clear that 'species' represents different, non-standard evolutionary entities across clades [8]. The boundaries between species and populations are opaque, as illustrated by the many species concepts in existence. If barcoding is to make a contribution to species identification and discovery, it must espouse a clear, modern concept of 'species' more specific than: 'Interpreting discontinuities in interspecific variation' [2]. Currently, by suggesting that phylogenetic background knowledge is not needed to identify 'gene species', barcoding is holding to a phenetic, essentialist, species concept [5,6]. The essentialist species concept is the notion that species have a 'true essence' and are definable as unchanging, discrete units. Essentialism contrasts with current evolutionary thinking that incorporates the complex, dynamic relationships seen between organisms and lineages in nature into our concept of species. Most species concepts implicitly include some idea of monophyly, or diagnosability based on shared derived traits and breeding patterns. Modern concepts of species are therefore broadly phylogenetic rather than essentialistic. Kress *et al.* do not venture into this territory, but use of any plant barcodes for species discovery will have to address species concepts, hopefully in a way that is more progressive than the current animal DNA barcoding efforts.

What should we be doing?

The hope of finding a single, short sequence of DNA from one gene that will reveal the identities of all plants or animals could be akin to a search for the Holy Grail. By virtue of the selective forces of evolution, there is unlikely to be a simple answer. Some genes, or parts of genes, might be better than others, and improved analyses are available that are not vulnerable to the pitfalls that make the distance-based searches currently in use undesirable. But the futility of trying to distill the identities of all members of all species from a short sequence of one gene will not change. A barcode might work for many species, but for a broad-scale analysis of life on Earth, even the most pragmatic person will agree that a success rate of 80% is unacceptable. It would leave millions of species misidentified and will not alleviate the taxonomic crisis of millions of species remaining undiscovered and undescribed.

Kress *et al.* [2] suggest that critics of barcoding are motivated by a desire to preserve morphological research; these critics are worried that barcodes will replace morphology and that such detractors are interested in preserving 'traditional' techniques for taxonomy and identification. This is, perhaps, an oversimplification. Many criticisms of barcoding are not made in the defense of morphological characters for their own sake [3,4], but rather emphasize the importance of multiple sources of data [7,9], including the use of multiple genes, and morphological and/or ecological characters in an analysis. Kress *et al.* appear to appreciate this necessity and explore the use of multiple markers. This commendable effort is congruent with those critics of barcoding who are not comparing the use of DNA sequences as a barcode against other single sources of characters, but rather the use of any single, uniformly inherited character set as a barcode against the necessary use of many data in an integrated approach. There are ways to enhance taxonomy, but there are no real shortcuts when dealing with a complex and contingent historical system.

Conclusions

Finding a minimum amount of gene sequence data that accurately represents the whole genome of all plants or animals is a daunting task. Kress *et al.* [2] have taken a positive step, but, in our view, the philosophical and practical barriers inherent in a single barcode make a successful conclusion to the search unlikely. Because none of the candidate markers worked for all samples, Kress *et al.* make the exciting suggestion that more than one locus is needed for plant barcode identifications. This is an important philosophical difference from the rest of the barcoding literature and commands careful consideration, because such an approach might overcome many of the pitfalls of traditional barcoding. Although never suggested overtly, Kress *et al.*'s paper might result in a barcoding 'compromise', in which the insufficiency of a small segment from a single gene is acknowledged, and multiple short segments from separate regions of the genome are used in compliment, hopefully in a phylogenetically robust context. Could this be a crossroads between integrative taxonomy and barcodes? More importantly, will it work?

References

1 Hebert, P.D.N. *et al.* (2003) Biological identifications through DNA barcodes. *Proc. R. Soc. B* 270, 313–321
2 Kress, J.W. *et al.* (2005) Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8369–8374
3 Prendini, L. (2005) Comment on 'Identifying spiders through DNA barcodes'. *Can. J. Zool.* 83, 481–491
4 Will, K.W. and Rubinoff, D. (2004) Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20, 47–55
5 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
6 Hogg, I.D. and Hebert, P.D.N. (2004) Biological identification of springtails (Collembola: Hexapoda) from the Canadian Arctic, using mitochondrial DNA barcodes. *Can. J. Zool.* 82, 749–754
7 Funk, D.J. and Omland, K.E. (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Syst.* 34, 397–423
8 Coyne, J.A. and Orr, H.A. (2004) *Speciation*, Sinauer
9 Rubinoff, D. and Holland, B.H. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Syst. Biol.* (in press)

0169-5347/$ - see front matter
doi:10.1016/j.tree.2005.10.019