

DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach

S.G. Newmaster, A.J. Fazekas, and S. Ragupathy

Abstract: DNA barcoding based on the mitochondrial cytochrome *c* oxidase I (*coxI*) sequence is being employed for diverse groups of animals with demonstrated success in species identification and new species discovery. Applying barcoding systems to land plants will be a more challenging task as plant genome substitution rates are considerably lower than those observed in animal mitochondria, suggesting that a much greater amount of sequence data from multiple loci will be required to barcode plants. In the absence of an obvious well-characterized plant locus that meets all the necessary criteria, a key first step will be identifying candidate regions with the most potential. To meet the challenges with land plants, we are proposing the adoption of a tiered approach wherein highly variable loci are nested under a core barcoding gene. Analysis of over 10 000 *rbcL* sequences from GenBank demonstrate that this locus could serve well as the core region, with sufficient variation to discriminate among species in approximately 85% of congeneric pair-wise comparisons. Use of a secondary locus can be implemented when required and can vary from group to group if necessary. The implementation of a barcoding tool has multiple academic and practical applications. It will speed routine identifications and the detection of alien species, advance ecological and taxonomic inquiry, permit fast and accurate forensic analysis of plant fragments, and can function as an additional layer of quality control in the food industry.

Key words: DNA barcoding, *rbcL*, land plants, species identification, taxonomy.

Résumé : On emploie le code à barre ADN basé sur une séquence du cytochrome oxydase *c* mitochondrial I (*coxI*) pour divers groupes d'animaux, avec un succès avéré pour l'identification des espèces et la découverte de nouvelles espèces. L'application du code barre aux plantes terrestres constituera un défi plus considérable, parce que les taux de substitution des génomes végétaux sont nettement plus faibles que ceux observés chez les mitochondries animales, ce qui suggère qu'il faudra utiliser une quantité beaucoup plus grande de données de séquences de multiples lieux, pour développer un code barre pour les plantes. En absence d'un lieu végétal évident et bien caractérisé qui rencontre tous les critères nécessaires, une première étape déterminante sera d'identifier des régions candidates les plus prometteuses. Pour rencontrer les défis des plantes terrestres, les auteurs proposent l'adoption d'une approche par étape où des lieux fortement variables se retrouvent dans un gène de code à barre central. L'analyse de plus de 10 000 séquences *rbcL* provenant de Genbank démontre l'utilité de ce lieu comme région centrale, avec suffisamment de variation pour discriminer parmi les espèces chez 85 % des comparaisons par paires de congénères. L'utilisation d'un second lieu peut être développée lorsque nécessaire et peut également varier d'un groupe à l'autre si nécessaire. Le développement d'un outil de code à barre s'ouvre sur de multiples applications académiques et pratiques. Il permettra d'accélérer les identifications usuelles et la détection d'espèces étrangères, les recherches poussées en écologie et taxonomie, les analyses médico-légales de fragments de végétaux, et pourra également servir d'élément additionnel pour le contrôle de qualité dans l'industrie alimentaire.

Mots clés : barre à code ADN, *rbcL*, plantes terrestres, identification des espèces, taxonomie.

[Traduit par la Rédaction]

Background

DNA barcoding as proposed by Hebert et al. (2003) is a system to aid species recognition and identification through

Received 22 December 2005. Published on the NRC Research Press Web site at <http://canjbot.nrc.ca> on 13 April 2006.

S.G. Newmaster,¹ A.J. Fazekas, and S. Ragupathy. Floristic Diversity Research Group, Biodiversity Institute of Ontario, Department of Integrative Biology, University of Guelph, Guelph, ON N1G 2W1, Canada.

¹Corresponding author (e-mail: snewmast@uoguelph.ca).

the characterization of a standard gene region across all organisms. The appeal and utility of the barcoding system is through the development of a comprehensive and rigorous database that is widely accessible. The system promises rapid and accurate identification and has demonstrated success in species discovery and identification from divergent animal taxa (e.g., Hebert et al. 2004a, 2004b; Barret and Hebert 2005; Monaghan et al. 2005), as well as red macroalgae (Saunders 2005).

Although there has been some criticism of the barcoding initiative, many concerns have resulted from a misunderstanding of its aims, or from misuse and misinterpretation

of terms (for a comprehensive response to critics of barcoding, see Hebert and Gregory (2005)). DNA barcoding is the use of a short gene sequence from a standardized region of the genome that can be used to help discover, characterize, and distinguish species, and to assign unidentified individuals to species (<http://www.barcodeoflife.org>; <http://www.barcoding.si.edu>). Barcoding has been equated with “molecular taxonomy” (Blaxter 2004) or the “DNA taxonomy” of Tautz et al. (2003), but these terms need to be separated. While one aim of barcoding is to quickly identify putative new species, this is a starting point for traditional taxonomic description. Molecular taxonomy on the other hand, stops at this point, only bracketing the “new species” with genetic distance measures. Such an approach using molecular operational taxonomic units (MOTU) (Floyd et al. 2002) based on genetic distance may be appropriate to define new cohesive taxonomic entities in particular taxa that are exceedingly small in size and lack obvious morphological variation, but for the majority of species, such an approach is inadequate and too narrow to form the basis for species descriptions.

The conflation of phylogenetics and barcoding is also prominent in much of the discussion surrounding barcoding. While sequence data can (and is) used to evaluate phylogenetic hypotheses, it is emphasized (Hebert and Gregory 2005) that the data produced for barcoding animals is meant to be used as an identification tool, and to aid in the discovery of new species.

Why do we need a molecular taxonomic tool?

An increasingly accepted view is that traditional taxonomic practices are insufficient on their own to cope with the growing need for accurate and accessible taxonomic information. Although approximately 1.7 million species have been described and named under the Linnaean system (Hawksworth and Kalin-Arroyo 1995), the total number of species on earth remains unknown, and estimates vary widely, ranging from 10 million to more than 100 million (May 1988; Hammond 1992; Hawksworth and Kalin-Arroyo 1995; Barcode of Life Database (<http://www.barcodinglife.com/>)). Even using conservative estimates, it is recognized that the number of species remaining to be discovered far outstrips the current resources of descriptive taxonomists and systematists (Godfray 2002; Blaxter 2004).

The task of recognizing new species has a certain urgency; the diversity of our biosphere is so large that the methodical cataloguing of new species by traditional methods is being outpaced by losses from human impacts. In the face of such mounting losses to biodiversity, the need to catalogue and describe life is greater than ever, and there is a growing realization that it will be critical to seek technological assistance for a species' initial recognition and its subsequent identification (Godfray 2002; Blaxter and Floyd 2003; Godfray and Knapp 2004). In these contexts, the standardized approach of DNA barcoding has great value.

Additionally, barcoding clearly has enormous potential to relieve taxonomists of routine identifications, providing more time to focus on new taxonomic hypotheses. Although it has been suggested that taxonomists are not expected to provide routine identifications (Will and Rubinoff 2004), the reality is that taxonomists are the most qualified to do

so, and therefore often receive such requests. Our laboratory accepts thousands of specimens annually from outside sources wanting expert identification (and unfortunately turns away even more). Probably 95% of these specimens fall into the “routine” category and occupy a corresponding amount of time. Rather than turning taxonomists into “a high tech service industry” (Lipscomb et al. 2003; Will and Rubinoff 2004), barcoding would relieve them of a “low tech”, time intensive service already provided. When multiplied across the thousands of taxonomists that provide such services, the increased amount of time available to focus on the 5% of rare, poorly characterized, and new species would greatly assist efforts in advancing taxonomy.

The ability to quickly put a name to an unknown specimen benefits not only conservationists, but is also a tremendous tool for ecologists as well (Janzen 2004, 2005). Small and hard to identify groups of organisms are often tacitly lumped together with the inference that they perform the same function in the ecosystem and occupy the same niche. Ecologists who do work with such groups are often hampered by the inability to identify saplings, roots, seeds, pollen, asexual stages, as well as other cohabitants of the ecosystem that function as food sources, habitat, predators, etc. This “taxonomic impediment” in ecology (Wheeler et al. 2004) is well served by a molecular approach. The use of barcoding will readily allow the identification of small plant fragments or sterile material, which previously would have been extraordinarily difficult or impossible to identify.

Finally, society-at-large is in need of routine diagnostic services for plants. Currently there is no routine identification service for cryptic species or difficult to identify pieces of plants such as leaves, seeds, pollen, roots, rhizomes, etc. Our laboratory is frequently requested to provide such identifications: invasive species propagules in quarantine at the border; the identity of unknown plant material in herbal products; and forensic identification of leaf fragments obtained by the Royal Canadian Mounted Police. It is these diverse needs that have prompted the development of barcoding tools and, once available, the applications will blossom.

Does a perfect plant barcoding locus exist?

Those involved in initiating efforts in plant barcoding have focussed on the search for a candidate locus for identifying species (Chase et al. 2005; Kress et al. 2005). These efforts are inspired by the success of the mitochondrial gene cytochrome *c* oxidase 1 (*cox1*) as the core of the global bio-identification system for animals (Hebert et al. 2003). An optimal barcoding locus for plants will naturally have similar characteristics: sufficient variation between species such that species level discrimination can be achieved, but minimal variation within species. Unlike barcoding in animals, however, the mitochondrial *cox1* gene is not a good candidate for land plants as plant mitochondrial genes typically exhibit lower nucleotide substitution rates than plastid or nuclear genes (Wolfe et al. 1987, 1989; Palmer and Herbon 1988; Laroche et al. 1995, 1997; Yang et al. 1999). The substitution rate in plastid genes, however, is also not great, with rates about one quarter the rate observed in nuclear

DNA and 10- to 20-fold less than mammalian mitochondrial DNA (Wolfe et al. 1987, 1989; Palmer and Herbon 1988).

Of the multitude of nuclear regions available, one potential barcode candidate is the nuclear ribosomal internal transcribed spacer region (ITS) (Stoeckle 2003; Chase et al. 2005; Kress et al. 2005), which is used extensively in systematic and phylogenetic treatments, often successfully discriminating species. Although it has been noted that there are a number of concerns regarding the use of ITS as a locus for phylogenetic inference (Alvarez and Wendel 2003), many of these issues are of less importance to barcoding. Perhaps the most problematic issue to barcoding is the presence of paralogous copies of ITS observed in some genera (e.g., Campbell et al. 2005). Other nuclear regions are being employed for phylogenetic analysis (e.g., Grob et al. 2004; Small et al. 2004), however, the paucity of known single copy genes that are also widespread across taxa (Mort and Crawford 2004) and the presence of multiple alleles at these sites preclude their use as the foundation of a barcoding system that is to be fast and efficient.

Exploratory work to establish the most promising locus for plant barcoding has therefore defaulted to the plastid genome along with nuclear ribosomal ITS (Chase et al. 2005; Kress et al. 2005). The reduced variation observed in chloroplasts has led to the natural conclusion to utilize multiple loci from the outset. It is well known that noncoding regions typically exhibit more variation than coding regions because of a presumed reduction of functional constraints (Gielly and Taberlet 1994; Shaw et al. 2005), and it would appear to be appropriate to narrow the search to these regions. Indeed, the recent work by Kress et al. (2005) has proposed the plastid *trnH-psbA* spacer as a suitable locus. This locus meets the criteria that these authors have proposed (in the particular species they examined) as necessary for a barcode locus: (i) significant species level genetic variability and divergence, (ii) an appropriately short sequence length so as to facilitate DNA extraction and amplification, and (iii) the presence of conserved flanking sites for developing universal primers. *trnH-psbA* does however fall short of an additional criterion suggested by Blaxter (2004), namely, ease of alignment and analysis. While ease of alignment is not a strictly necessary criterion for DNA identification, it is a critical requirement for developing bioinformatics tools. For example, Kress et al. (2005) reported ITS and *trnH-psbA* sequences that were alignable within genera, but problematic above that rank. The presence of multiple indels that overlap (as in *trnH-psbA*) makes homology assessment and therefore accurate alignment difficult or impossible (Hamilton et al. 2003; Yamashiro et al. 2004).

Bioinformatics plays a critical role in barcoding. Unlike typical phylogenetic analyses, which may have a few hundred individuals in the data set, the aim of the barcode initiative is to include all species, with multiple representatives from each. For land plants, this will mean approximately 2–3 million individual specimens. The importance of bioinformatics tools to the system as a whole therefore cannot be underestimated, and an essential part of the process as an efficient and accurate alignment algorithm. Alignment methods such as Hidden Markov Models, currently used in the Barcode of Life Database, are founded on information derived from the analysis of significant secondary structure,

which is reduced or absent in noncoding DNA. Thus the needs of the bioinformatics process (a coding region) versus the desire to use a maximally variable locus (noncoding DNA) are in conflict. Further complicating the utility of noncoding loci is the fact that structural rearrangements in the chloroplast genome have eliminated some spacer regions; of the 21 noncoding plastid regions examined by Shaw et al. (2005) the three most phylogenetically informative were unavailable in some lineages.

Even if the “perfect” plant barcoding locus were to be found, it is recognized that reliance on a single (usually) maternally inherited gene will be problematic in groups that exhibit hybridization and introgression. In species complexes that exhibit extensive introgression, incorporation of multiple nuclear regions will be a necessity to make confident identifications (Chase et al. 2005). The suggestion to include multiple loci (Kress et al. 2005) in barcoding systems was welcomed by critics of barcoding (Rubinoff et al. 2006). While the use of multiple loci is a straightforward response to the challenges for barcoding plants, the system must retain a minimal complement of loci for it to remain a fast and efficient tool.

A tiered approach

Given the issues above, it seems straightforward therefore, that the adoption of an initial approach including both coding and noncoding regions will be optimal. Instead of simply adding another locus to increase the string of bases, a tiered method can be created in which a first tier coding region common across the land plants provides resolution at a certain rank (e.g., family or genus) and a more variable (coding or noncoding) region provides resolution to the species level. Difficulty in aligning noncoding regions from highly divergent genera would become less of an issue, as alignments at this second tier would only need to occur for small numbers of taxa nested together under a common first tier sequence. By anchoring the system with a universal standard coding region, various and multiple noncoding regions can be used, the choice of which would depend on the group of interest. The flexibility in choice of a second tier locus will be an enormous advantage given the variation observed in substitution rates. This approach will also have the advantage of keeping the complement of regions needed for any particular group at two, preserving the current efficiency of the system.

Evaluation of *rbcL* as a core gene

One obvious choice for evaluation as a potential standard core coding region is *rbcL*, given its universality and ease of amplification and alignment. Although it has been suggested (Kress et al. 2005) that *rbcL* has been previously rejected as a locus for species identification by Gielly and Taberlet (1994), Renner (1999), and Salazar et al. (2003), we were interested to know whether it would serve as a first tier locus as our own initial cursory analysis indicated that it showed promise. We chose to focus on *rbcL* because it is the most characterized plastid coding region in GenBank, with wide representation from all major groups, and will therefore provide a good baseline for comparison with other plastid genes.

Table 1. Number of genera and species used to calculate congeneric pair-wise distance estimates and the range of genetic distance estimates for six plant categories, with a count of species having at least one zero distance estimate in congeneric comparisons.

	No. of genera	No. of species	Range of genetic distance estimates (%)	Total no. of congeneric pair-wise comparisons	No. of species with a genetic distance estimate of zero in congeneric comparisons	% of total species
Bryophytes	126	499	0–18	1 802	16	3.2
Ferns and allies	108	988	0–33	29 741	116	11.7
Gymnosperms	38	406	0–8	17 536	196	48.3
Paleoherbs, magnoliids, and monocots	345	1501	0–14	11 247	233	15.5
Rosids	292	1124	0–10	7 646	176	15.7
Asterids	243	1026	0–19	12 120	40	3.9
Total	1152	5544	0–33	80 092	779	14.1

Recently, Chase et al. (2005) reiterated the fact that the utility of a particular gene for phylogenetic analysis does not equate with its utility for barcoding. These authors further suggest that a first task is to evaluate regions that have been previously used for phylogenetics, in a barcoding context. Using data from GenBank, they used the BLAST algorithm to test the ability of *rbcL* as well as various combinations of the nuclear ribosomal ITS and 5.8S regions to make accurate identifications. For *rbcL* at the species level, the percentage of incorrect BLAST hits with a higher score than the last correct hit was 16.95%. At the genus level, the percentage of species assigned to an incorrect genus before the lowest ranked correct assignment was 67.71%.

Rather than using the approach of Chase et al. (2005), we used a distance-based method, as that is the method currently employed in the Barcode of Life Database for making identifications or suggesting a closest match. Based on available data from GenBank, we evaluated the potential for *rbcL* to form the “backbone” of a two-tiered approach to the molecular identification of plants. Specifically, we were interested to know how well *rbcL* would resolve congeneric species. Over 17 000 *rbcL* sequences from the land plants were deposited in GenBank as of August 2005. The length of the *rbcL* exon is approximately 1428 bp, however, many *rbcL* sequences in GenBank are of short length, so we therefore limited the inclusion of sequences in our analysis to those that were at least 1000 bp in length. Using the approximately 10 300 sequences remaining, we aligned each of six categories, which correspond to broad plant groups: bryophytes; ferns and allies; gymnosperms; paleoherbs, monocots, and magnoliids; rosids; and asterids. Using the aligned datasets, we then calculated pair-wise distances among congeneric species using the Kimura 2-parameter distance model for each of the categories. Although the sequences used had greater than 1000 bases, a small number of them (approx. 50) had an excessive amount of bases that were indeterminate. After the calculation of distance estimates, we identified and removed estimates from species pairs that had <600 bp overlap.

The distance estimates that we calculated had broad ranges, from 0% to a high of 33% in the ferns and allies (Table 1). The high rates observed in this group are from congeneric comparisons between species of *Selaginella*,

some of which exhibit extreme substitution rates (Korall and Kenrick 2002). To explore the distribution of variation in the distance estimates we created a histogram showing the percentage of congeneric pair-wise comparisons that fall into classes of genetic distance with intervals of 1% (Fig. 1). These distributions indicate that for *rbcL*, most genetic distances values between congeneric species are over 1%.

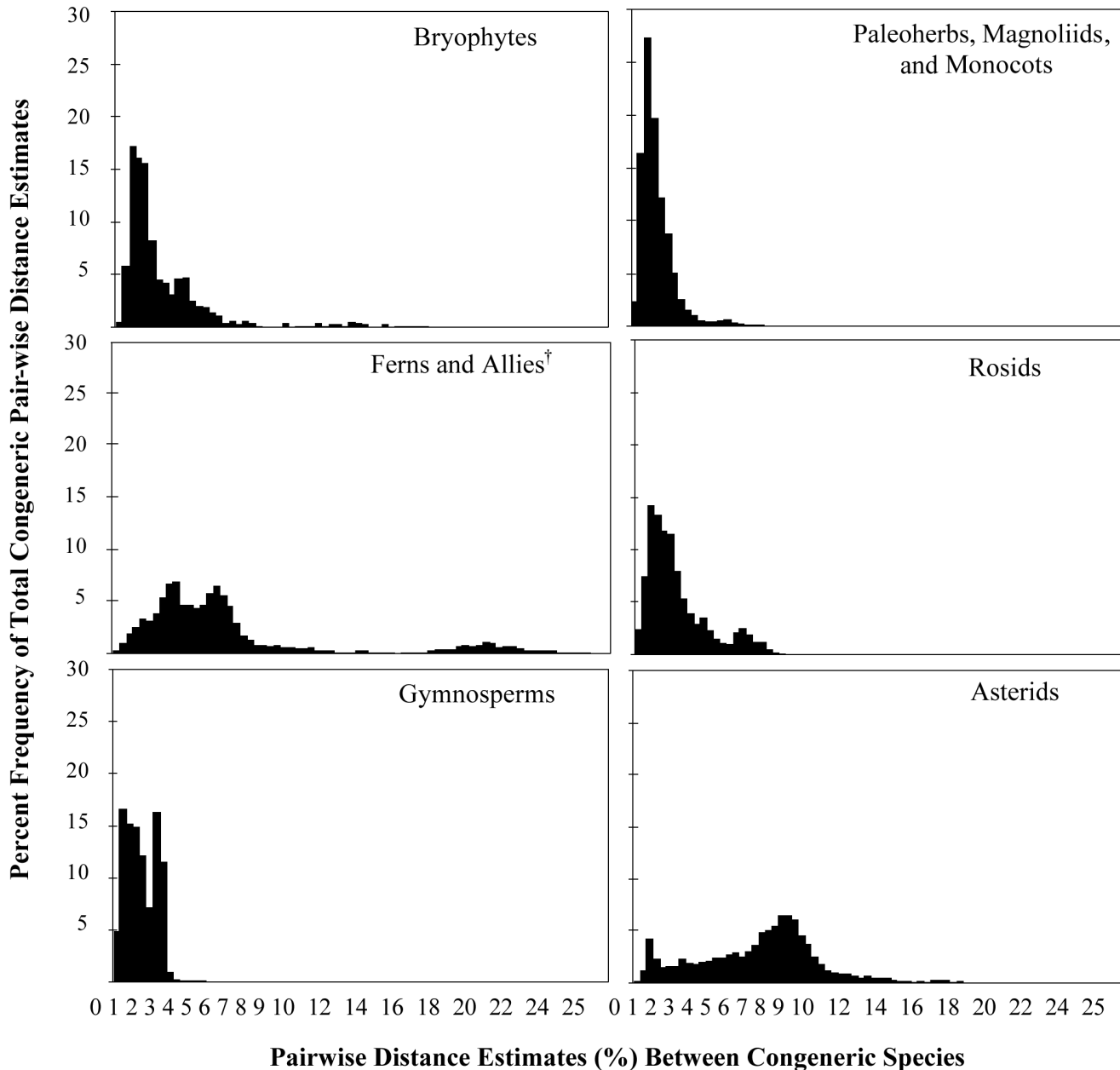
A count of the number of sequences that had a zero distance value from at least one other sequence is given in Table 1 and shows considerable variation in the ability of *rbcL* to resolve congeners. Notably, the gymnosperms have a very high failure rate, which is primarily the result of the disproportionate representation of the genus *Pinus* (over 25%) in this category, whose species do not resolve well with *rbcL*.

While *rbcL* seems to be a reasonable candidate for a first tier barcoding locus, there are other regions that may prove to be of greater utility. The plastid gene *matK*, for example, has a substitution rate that is 2–3 times greater than *rbcL* in angiosperms, however, there is only a small amount of data available for the bryophytes or ferns, precluding a quick evaluation. Given the deep divergences in these groups observed with *rbcL*, one might expect *matK* to be even more useful than *rbcL* as a barcoding locus. In other groups, *matK* seems to have little additional variation; in the genus *Pinus*, for which many species displayed identical *rbcL* sequences, *matK* is also insufficiently variable to resolve species (Gernandt et al. 2005). Clearly any potential first tier candidate locus will need to be widely tested before it is adopted, and it is unlikely that any single candidate gene will prove to be “perfect” in all respects or work equally well for barcoding in all plant groups.

The simple addition of a second locus commonly applied across all taxa may not add further resolution in some cases. Genes that are common across all land plants will likely be too conserved in some groups to be of utility to barcoding. Highly variable gene regions such as those identified by Shaw et al. (2005) may be missing in some groups of plants. This is our rationale for suggesting a tiered approach, where the second tier is taxon specific. At the second tier, we expect that only a handful of different regions will be necessary to provide species level resolution for all plants.

Our analysis suggests that *rbcL* would perform well as the core locus in a multigene, tiered approach to barcoding in the land plants. Ultimately, other more quickly evolving

Fig. 1. Frequency histogram of congeneric pair-wise distances, estimated using the Kimura 2-parameter substitution rate, for six plant categories based on *rbcL* data obtained from GenBank. The interval on the *x*-axis is 1%. †, ferns and allies had 210 pair-wise estimates (0.7% of the total), distributed from 25% to 33%, that are not shown on the histogram.



plastid regions may prove to be a better choice, yielding more information in a shorter read length. When coupled with ITS or a plastid noncoding region, an approach that fits well with the bioinformatics process can be formed.

Although the sequence data available for *rbcL* is large in terms of the number of species represented, we acknowledge that the data set does have some drawbacks: species coverage is uneven and variable, many taxa are misidentified, sequence quality is unknown, and it is uncommon to have more than one individual per species represented. The inclusion of multiple representatives for each species is an important component of any barcoding effort. Characterization of intraspecific variation and detection of cryptic species can-

not be accomplished with representation from only a single individual (e.g., Hebert et al. 2004b). To provide a comprehensive evaluation for any locus, however, it would effectively mean sequencing a large percentage of species, a de facto realization of the goal. This “unknown factor” is an issue with any discovery-based project, but as Hebert and Gregory (2005) observe, such projects always produce numerous unanticipated hypotheses. In the course of analysing available GenBank sequences, we have observed a number of patterns that merit further exploration.

Based on the results from our analysis of *rbcL*, we have initiated a barcoding project using *rbcL* for the mosses of Ontario. Mosses exhibit a number of characteristics that il-

illustrate the appeal of barcoding. They are a diverse and cryptic group of organisms that are small in size and taxonomically challenging. The consequent lack of expertise in identification has limited the ability to study the ecology of this group. Although we have only sampled a small portion of the Ontario mosses (7.2% of species, 14.1% of genera) our preliminary results are promising: of 114 moss samples (39 species, across 30 genera), *rbcL* demonstrates a small amount of variation within species (mean = 0.21%) and between species distance estimates range from 0.79% to 13.89%. This dataset is currently limited to Ontario, and includes a number of circumtemperate species. The degree of additional variation that will be observed when samples are included from a broader range is unknown; therefore we are expanding our moss dataset to include populations from the entire global range of each species. On a practical level, an unknown specimen is usually from a known geographic location, so it is not necessary to characterize variation from the entire range for a barcoding system to provide utility at a regional level. With this small dataset, we are interested in whether or not a particular species concept can be diagnosed using *rbcL* for different populations of mosses separated by great distances (>1000 km) across Ontario. In addition to *rbcL*, we are surveying a number of other gene regions to determine which might be most appropriate as a first tier locus.

Conclusion

Establishing a standard barcoding region(s) for land plants is not as obvious a process as was the selection of *cox1* for animals. From the volume of data available, and current understanding of genome structure, nucleotide substitution rates, and variation in rate distribution, the likelihood of finding a single gene for the plants that is as easy to use, and works as well as *cox1* does for animal barcoding, is slim.

It is straightforward that a multigene approach for the plant kingdom will be necessary. The use of multiple loci, however, will frustrate comparisons between taxa, without the inclusion of a standard common region. The inclusion of a coding region common in land plants, such as *rbcL*, can serve as a baseline for comparison, and in a tiered approach, allows flexibility in choice of a secondary locus.

In this respect, *rbcL* while not “perfect” meets most of the desired attributes for a first tier barcoding locus, and the data we have evaluated indicate that it can be used to resolve congeneric species in approximately 85% of cases. In fact, our results suggest it may serve well as the sole locus for barcoding mosses, however, evaluation of other plastid coding regions may yet reveal a better candidate for all land plants. In other groups some of the plastid noncoding regions evaluated by Shaw et al. (2005) or *trnH-psbA* proposed by Kress et al. (2005) appear promising as a second tier locus.

Given the pace of advancement in technology, it is not unrealistic that in the span of a few years we may be using a barcoding tool for routine identifications, discovering new species, solving ecological puzzles, controlling the pathways of invasive species, and for quality control in the food and herbal industries. Certainly in the course of development of

a rigorous and comprehensive database multiple new ecological and taxonomic hypotheses will emerge.

Acknowledgements

This research was supported by Genome Canada and the Canadian Foundation for Innovation. We thank Paul Hebert and Jeremy deWaard for their guidance and wisdom, and Sujeevan Ratnasingham for his assistance with the numerical analysis of this large dataset. We thank Heather Cole, Melissa Berry, and Julian Kwan for their assistance in our laboratory at the Biodiversity Institute Herbarium. We would also like to thank Robert Hanner and Ryan Gregory for reviewing an earlier version of the manuscript, and Randall Bayer and Jean Bousquet for their formal review.

References

- Alvarez, I., and Wendel, J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Mol. Phylogenet. Evol.* **29**: 417–434. PMID: 14615184.
- Barret, D.H., and Hebert, P.D.N. 2005. Identifying spiders through DNA barcodes. *Can. J. Zool.* **83**: 481–491.
- Blaxter, M. 2004. The Promise of a DNA taxonomy. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**: 669–679. PMID: 15253352.
- Blaxter, M., and Floyd, R. 2003. Molecular taxonomics for biodiversity surveys: already a reality. *Trends Ecol. Evol.* **18**: 268–269. doi:10.1016/S0169-5347(03)00102-2.
- Campbell, C.S., Wright, W.A., Cox, M., Vining, T.F., Major, C.S., and Arsenault, M.P. 2005. Nuclear ribosomal DNA internal transcribed spacer 1 (ITS1) in *Picea* (Pinaceae): sequence divergence and structure. *Mol. Phylogenet. Evol.* **35**: 165–185. PMID: 15737589.
- Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesana-kurthi, R.P., Haidar, N., and Savolainen, V. 2005. Land plants and DNA barcodes: short-term and long-term goals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**: 1889–1895. PMID: 16214746.
- Floyd, R., Eyualem, A., Papert, A., and Blaxter, M.L. 2002. Molecular barcodes for soil nematode identification. *Mol. Ecol.* **11**: 839–850. doi:10.1046/j.1365-294X.2002.01485.x. PMID: 11972769.
- Gernandt, D.S., López, G.G., García, S.O., and Liston, A. 2005. Phylogeny and classification of *Pinus*. *Taxon*, **54**: 29–42.
- Gielly, L., and Taberlet, P. 1994. The use of chloroplast DNA to resolve plant phylogeny: noncoding versus *rbcL* sequences. *Mol. Biol. Evol.* **11**: 769–777. PMID: 7968490.
- Godfray, H.C.J. 2002. Challenges for taxonomy. *Nature* (London), **417**: 17–19. doi:10.1038/417017a. PMID: 11986643.
- Godfray, H.C.J., and Knapp, S. 2004. Introduction. *In* Taxonomy for the twenty-first century—Special Issue. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**: 559–569. PMID: 15253344.
- Grob, G.B.J., Gravendeel, B., and Eurlings, M.C.M. 2004. Potential phylogenetic utility of the nuclear FLORI-CAULA/LEAFY second intron: comparison with three chloroplast DNA regions in *Amorphophallus* (Araceae). *Mol. Phylogenet. Evol.* **30**: 13–23. PMID: 15022754.
- Hamilton, M.B., Braverman, J.M., and Soria-Hernanz, D.F. 2003. Patterns and relative rates of nucleotide and insertion/deletion evolution at six chloroplast intergenic regions in New World species of the Lecythidaceae. *Mol. Biol. Evol.* **20**: 1710–1721. PMID: 12832633.
- Hammond, P. 1992. Species inventory. *In* Global biodiversity: status of the earth's living resources. Edited by B. Groombridge. Chapman and Hall, London. pp. 17–39.

- Hawksworth, D.L., and Kalin-Arroyo, M.T. 1995. Magnitude and distribution of biodiversity. In *Global biodiversity assessment*. Edited by V.H. Heywood. Cambridge University Press, Cambridge, UK. pp. 107–191.
- Hebert, P.D.N., and Gregory, T.R. 2005. The promise of DNA barcoding for taxonomy. *Syst. Biol.* **54**: 852–859. doi:10.1080/10635150500354886. PMID: 16243770.
- Hebert, P.D.N., Cywinka, A., Ball, S.L., and deWaard, J.R. 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B. Biol. Sci.* **270**: 313–321.
- Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., and Francis, C.M. 2004a. Identification of birds through DNA barcodes. *PLoS Biol.* **2**: e312. doi:10.1371/journal.pbio.0020312. PMID: 15455034.
- Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., and Hallwachs, W. 2004b. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrartes fulgerator*. *Proc. Natl. Acad. Sci. USA*, **101**: 14812–14817. doi:10.1073/pnas.0406166101. PMID: 15465915.
- Janzen, D.H. 2004. Now is the time. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**: 731–732. PMID: 15253359.
- Janzen, D.H. 2005. How to conserve wild plants? Give the world the power to read them. *Foreword to Plant conservation: A natural history approach*. Edited by G.A. Krupnick and J. Kress. Chicago University Press, Chicago, Ill. pp. ix–xii.
- Korall, P., and Kenrick, P. 2002. Phylogenetic relationships in selaginellaceae based on *rbcL* sequences. *Am. J. Bot.* **89**: 506–517.
- Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A., and Janzen, D.H. 2005. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA*, **102**: 8369–8374. doi:10.1073/pnas.0503123102. PMID: 15928076.
- Laroche, J., Li, P., and Bousquet, J. 1995. Mitochondrial DNA and monocot-dicot divergence time. *Mol. Biol. Evol.* **12**: 1151–1156.
- Laroche, J., Li, P., Maggia, L., and Bousquet, J. 1997. Molecular evolution of angiosperm mitochondrial introns and exons. *Proc. Natl. Acad. Sci. USA*, **94**: 5722–5727. doi:10.1073/pnas.94.11.5722. PMID: 9159140.
- Lipscomb, D., Platnik, N., and Wheeler, Q. 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends Ecol. Evol.* **18**: 65–66. doi:10.1016/S0169-5347(02)00060-5.
- May, R.M. 1988. How many species are there on earth? *Science* (Washington, D.C.), **241**: 1441–1449.
- Monaghan, M.T., Balke, M., Gregory, T.R., and Vogler, A.P. 2005. DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**: 1925–1933. PMID: 16214750.
- Mort, M.E., and Crawford, D.J. 2004. The continuing search: low-copy nuclear sequences for lower-level plant molecular phylogenetic studies. *Taxon*, **52**: 257–261.
- Palmer, J.D., and Herbon, L.A. 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J. Mol. Evol.* **28**: 87–97. doi:10.1007/BF02143500. PMID: 3148746.
- Renner, S.S. 1999. Circumscription and phylogeny of the Laurales: Evidence from molecular and morphological data. *Am. J. Bot.* **86**: 1301–1315. PMID: 10487818.
- Rubinoff, D., Cameron, S., and Will, K. 2006. Are plant DNA barcodes a search for the Holy Grail? *Trends Ecol. Evol.* **21**: 1–2. doi:10.1016/j.tree.2005.10.019.
- Salazar, G.A., Chase, M.W., Arenas, M.A.S., and Ingrouille, M. 2003. Phylogenetics of Cranichideae with emphasis on Spiranthinae (Orchidaceae, Orchidoideae): evidence from plastid and nuclear DNA sequences. *Am. J. Bot.* **90**: 777–795.
- Saunders, G.W. 2005. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**: 1879–1888. doi:10.1098/rstb.2005.1719. PMID: 16214745.
- Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.S., Liu, W., Miller, J., Chaw, S.K., Winder, C.T., Schilling, E.E., and Small, R.L. 2005. The tortoise and the hare: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* **92**: 142–166.
- Small, R.L., Cronn, R.C., and Wendel, J.F. 2004. Use of nuclear genes for phylogeny reconstruction in plants. *Aust. Syst. Bot.* **17**: 145–170. doi:10.1071/SB03015.
- Stoeckle, M. 2003. Taxonomy, DNA, and the bar code of life. *Bioscience*, **53**: 796–797.
- Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., and Vogler, A.P. 2003. A plea for DNA taxonomy. *Trends Ecol. Evol.* **18**: 70–74. doi:10.1016/S0169-5347(02)00041-1.
- Wheeler, Q.D., Raven, P.H., and Wilson, E.O. 2004. Taxonomy: impediment or expedient? *Science* (Washington, D.C.), **303**: 285. doi:10.1126/science.303.5656.285. PMID: 14726557.
- 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. doi:10.1111/j.1096-0031.2003.00008.x.
- Wolfe, K.H., Li, W.H., and Sharpe, P.M. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA*, **84**: 9054–9058. PMID: 3480529.
- Wolfe, K.H., Sharp, P.M., and Li, W.H. 1989. Rates of synonymous substitution in plant nuclear genes. *J. Mol. Evol.* **29**: 208–211. doi:10.1007/BF02100204.
- Yamashiro, T., Fukuda, T., Yokoyama, J., and Maki, M. 2004. Molecular phylogeny of *Vincetoxicum* (Apocynaceae-Asclepiadoideae) based on the nucleotide sequences of cpDNA and nrDNA. *Mol. Phylogenet. Evol.* **31**: 689–700. PMID: 15062803.
- Yang, Y.W., Lai, K.N., and Li, W.H. 1999. Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. *J. Mol. Evol.* **48**: 597–604. PMID: 10198125.