A DNA barcode for land plants

CBOL Plant Working Group[1]

Communicated by Daniel H. Janzen, University of Pennsylvania, Philadelphia, PA, May 27, 2009 (received for review March 18, 2009)

DNA barcoding involves sequencing a standard region of DNA as a tool for species identification. However, there has been no agreement on which region(s) should be used for barcoding land plants. To provide a community recommendation on a standard plant barcode, we have compared the performance of 7 leading candidate plastid DNA regions (*atpF–atpH* spacer, *matK* gene, *rbcL* gene, *rpoB* gene, *rpoC1* gene, *psbK–psbI* spacer, and *trnH–psbA* spacer). Based on assessments of recoverability, sequence quality, and levels of species discrimination, we recommend the 2-locus combination of *rbcL*+*matK* as the plant barcode. This core 2-locus barcode will provide a universal framework for the routine use of DNA sequence data to identify specimens and contribute toward the discovery of overlooked species of land plants.

matK | *rbcL* | species identification

Large-scale standardized sequencing of the mitochondrial gene *CO1* has made DNA barcoding an efficient species identification tool in many animal groups (1). In plants, however, low substitution rates of mitochondrial DNA have led to the search for alternative barcoding regions. From initial investigations of plastid regions (2–4), 7 leading candidates have emerged (5, 6). Four are portions of coding genes (*matK*, *rbcL*, *rpoB*, and *rpoC1*), and 3 are noncoding spacers (*atpF–atpH*, *trnH–psbA*, and *psbK–psbI*). Different research groups have proposed various combinations of these loci as their preferred plant barcodes, but no consensus has emerged (5–12). This lack of an agreed standard has impeded progress in plant barcoding.

Our aim here is to identify a standard DNA barcode for land plants. To achieve this goal, we have pooled data across laboratories including sequence data from 907 samples, representing 445 angiosperm, 38 gymnosperm, and 67 cryptogam species. Using various subsets of these data, we evaluated the 7 candidate loci using criteria in the Consortium for the Barcode of Life's (CBOL) data standards and guidelines for locus selection (http://www.barcoding.si.edu/protocols.html). *Universality*: Which loci can be routinely sequenced across the land plants? *Sequence quality and coverage*: Which loci are most amenable to the production of bidirectional sequences with few or no ambiguous base calls? *Discrimination*: Which loci enable most species to be distinguished?

Results

Universality. Direct universality assessments using a single primer pair for each locus in angiosperms resulted in 90%–98% PCR and sequencing success for 6/7 regions. Success for the seventh region, *psbK–psbI*, was 77% (Fig. 1*A*). Greater problems were encountered in other land plant groups, with *rpoB*, *matK*, *atpF–atpH*, and *psbK–psbI* all showing <50% success in gymnosperms and/or cryptogams based on data compiled from several laboratories (Fig. 1*A*).

Sequence Quality. Evaluation of sequence quality and coverage from the candidate loci demonstrated that high quality bidirectional sequences were routinely obtained from *rbcL*, *rpoC1*, and *rpoB* (Fig. 1*B*, *x* axis). The remaining 4 loci required more manual editing and produced fewer bidirectional reads. *matK* performed best of this group, although it showed discordance between forward and reverse reads more frequently than other coding regions. The greatest problems in obtaining bidirectional sequences with few ambiguous bases were encountered with the intergenic spacers *trnH–psbA* and *psbK–psbI*, in part attributable to a high frequency of mononucleotide repeats disrupting individual sequencing reads.

Species Discrimination. Among 397 samples successfully sequenced for all 7 loci, species discrimination for single-locus barcodes ranged from 43% (*rpoC1*) to 68%–69% (*psbK–psbI* and *trnH–psbA*), with *rbcL* and *matK* providing 61% and 66% discrimination respectively (rank order: *rpoC1*<*rpoB*<*atpF–*

Author contributions: P.M.H., L.L.F., J.L.S., M.H., S.R., M.v.d.B., M.W.C., R.S.C., D.L.E., A.J.F., S.W.G., K.E.J., K.-J.K., W.J.K., H.S., S.C.H.B., C.v.d.B., M.C., T.A.J.H., B.C.H., G.P., J.E.R., G.A.S., V.S., O.S., M.J.W., and D.P.L. designed research; D.L.E., A.J.F., K.E.J., J.v.A.S., D.B., K.S.B., K.M.C., J.C., A.C., J.J.C., F.C., D.S.D., C.S.F., M.L.H., L.J.K., P.R.K., J.S.K., Y.D.K., R.L., H.-L.L., D.G.L., S.M., O.M., I.M., S.G.N., C.-W.P., D.M.P., and D.-K.Y. performed research; L.L.F., J.L.S., M.H., S.R., and D.P.L. analyzed data; and P.M.H., S.W.G., S.C.H.B., and D.P.L. wrote the paper.

Conflict of interest statement: Following the publication of Lahaye et al. (PNAS 105:2923, 2008), the process of filing a patent on DNA barcoding of land plants using *matK* was initiated by V.S., M.v.d.B., R.L., and D.B., but because of the lack of commercial interest the patent application was subsequently dropped.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database. For a list of accession numbers, see SI Table 1. FASTA files of sequences are available on request.

See Commentary on page 12569.

[1]CBOL Plant Working Group: Peter M. Hollingsworth[a,2], Laura L. Forrest[a], John L. Spouge[b], Mehrdad Hajibabaei[c], Sujeevan Ratnasingham[c], Michelle van der Bank[d], Mark W. Chase[e], Robyn S. Cowan[e], David L. Erickson[f], Aron J. Fazekas[g], Sean W. Graham[h], Karen E. James[i], Ki-Joong Kim[j], W. John Kress[f], Harald Schneider[i], Jonathan van AlphenStahl[e], Spencer C.H. Barrett[k], Cassio van den Berg[l], Diego Bogarin[m], Kevin S. Burgess[k,n], Kenneth M. Cameron[o], Mark Carine[i], Juliana Chacón[p], Alexandra Clark[a], James J. Clarkson[e], Ferozah Conrad[q], Dion S. Devey[e], Caroline S. Ford[r], Terry A.J. Hedderson[s], Michelle L. Hollingsworth[a], Brian C. Husband[g], Laura J. Kelly[a,e], Prasad R. Kesanakurti[g], Jung Sung Kim[j], Young-Dong Kim[t], Renaud Lahaye[d], Hae-Lim Lee[i], David G. Long[a], Santiago Madriñán[p], Olivier Maurin[d], Isabelle Meusnier[c], Steven G. Newmaster[g], Chong-Wook Park[u], Diana M. Percy[h], Gitte Petersen[v], James E. Richardson[a], Gerardo A. Salazar[w], Vincent Savolainen[e,x], Ole Seberg[v], Michael J. Wilkinson[r], Dong-Keun Yi[j], and Damon P. Little[y]

[a]Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, United Kingdom; [b]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Computational Biology Branch, Bethesda, MD 20894; [c]Biodiversity Institute of Ontario, Department of Integrative Biology, University of Guelph, Guelph, ON, Canada N1G 2W1; [d]Department of Botany and Plant Biotechnology, University of Johannesburg, P.O. Box 524, Auckland Park, Johannesburg 2006, South Africa; [e]Royal Botanic Gardens, Kew, Richmond TW9 3DS, United Kingdom; [f]Department of Botany, Smithsonian Institution, Washington DC, 20013-7012; [g]Department of Integrative Biology, University of Guelph, Guelph, ON, Canada N1G 2W1; [h]UBC Botanical Garden and Centre for Plant Research, Faculty of Land and Food Systems, and Department of Botany, University of British Columbia, Vancouver, BC, Canada V6T 1Z4; [i]Botany Department, Natural History Museum, London SW7 5BD, United Kingdom; [j]School of Life Sciences and Biotechnology, Korea University, Seoul 136-701, Korea; [k]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada M5S 3B2; [l]Laboratório de Sistemática Molecular de Plantas, Universidade Estadual de Feira de Santana, Departamento de Ciências Biológicas, 44031-460, Feira de Santana, Bahia, Brazil; [m]Jardín Botánico Lankester, Universidad de Costa Rica, Cartago, Costa Rica; [n]Department of Biology, Columbus State University, Columbus, GA 31907-5645; [o]Department of Botany, University of Wisconsin, Madison, WI 53508; [p]Universidad de los Andes, Apartado Aéreo 4976, Bogotá, D.C., Colombia; [q]Leslie Hill Molecular Systematics Laboratory, SANBI, Kirstenbosch Research Centre, Claremont 7735, Cape Town, South Africa; [r]Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Ceredigion SY23 3DA, United Kingdom; [s]Department of Botany, University of Cape Town, Rondebosch 7700, South Africa; [t]Department of Life Sciences, Hallym University, Chuncheon 200-702, Korea; [u]School of Biological Sciences, Seoul National University, Seoul 151-742, Korea; [v]Natural History Museum of Denmark, University of Copenhagen, 1307 Copenhagen K, Denmark; [w]Instituto de Biología, Universidad Nacional Autónoma de México, 04510 México, D.F., Mexico; [x]Imperial College London, Silwood Park Campus, Ascot SL5 7PY, United Kingdom; and [y]Cullman Program for Molecular Systematics, New York Botanical Garden, Bronx, NY, 10458-5126

[2]To whom correspondence should be addressed. E-mail: P.Hollingsworth@rbge.org.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0905845106/DCSupplemental.

Downloaded from https://www.pnas.org by 205.139.151.11 on August 6, 2025 from IP address 205.139.151.11.

Tecan Ex. 2116, Page 1 of 4
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

SEE COMMENTARY

ECOLOGY

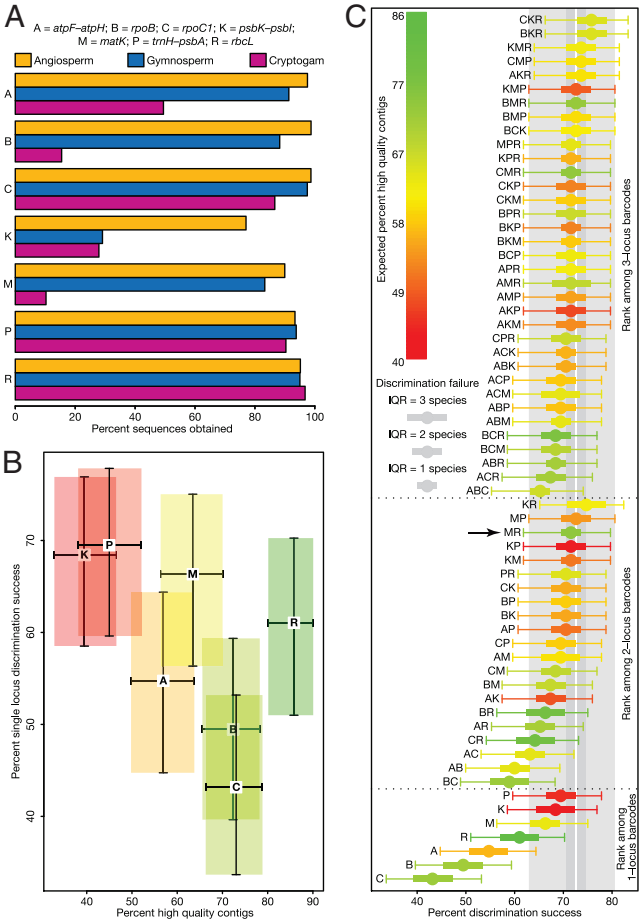

Fig. 1. Comparison of the performance of 7 candidate barcoding loci (see locus codes at head of Fig. 1*A*). (*A*) Universality success based on 170 angiosperm samples compared under similar conditions, and community-wide data for up to 81 gymnosperm and 156 cryptogam samples. (*B*) Assessment of sequence quality calculated as the percentage of 190 seed plant samples from which high quality bidirectional sequences (contigs) could be assembled (see *Materials and Methods* for trace-quality criteria), plotted against the percentage species discrimination for single-locus barcodes. 95% confidence intervals are indicated. Colors reflect sequence quality (red, worse; green, better). (*C*) Discrimination success for 1–3 and 7 locus barcodes for species for which multiple individuals from multiple congeneric species were sampled, and all 7 loci were recovered. Outer error bars (thin lines) demarcate 95% confidence intervals. Inner error bars (thick lines) indicate the relative magnitude of discrimination failure as measured by the interquartile range (IQR) for the number of species that are indistinguishable from a given query sequence. Discrimination success from all 7 loci is shown with a white line, with the associated 95% confidence interval in light gray, and the magnitude of discrimination failure in dark gray. Colors indicate the average percentage of finished bidirectional sequences expected for each locus combination. The arrow indicates the recommended standard 2-locus barcode.

atpH<*rbcL*<*matK*<*psbK–psbI*<*trnH–psbA*; Fig. 1*B*, *y* axis). Two-locus combinations gave 59%–75% resolution, and 3-locus combinations 65%–76% (Fig. 1*C*). Ten of the 2-locus combinations gave 70%–75% discrimination. The top 5 of these involved various combinations of *rbcL*, *psbK–psbI*, *matK*, and *trnH–psbA*. Using all 7 loci, 73% of species were discriminated. When the species discrimination analyses are extended to the full sample, which includes those that failed to sequence for 1 or more loci, the rank order among single-locus comparisons is *rpoC1* (38%), *rpoB* (40%), *atpF–atpH* (50%), *matK* (57%), *rbcL* (58%), *trnH–psbA* (58%), and *psbK–psbI* (64%). The rise in relative performance of *rbcL* is associated with its strong (87%) discriminatory power in the cryptogam samples. These were excluded from the preceding analyses as all had missing data from 1 or more loci.

Discussion

An ideal DNA barcode should be routinely retrievable with a single primer pair, be amenable to bidirectional sequencing with little requirement for manual editing of sequence traces, and provide maximal discrimination among species. Based on these criteria, 4 of the candidate loci can be excluded (Fig. 1 *A* and *B*). Both *rpoC1* and *rpoB* performed well in terms of universality and/or sequence quality, but had low discriminatory power; *atpF–atpH* fell below the median for species resolution in single and multilocus barcodes and for recovery of high-quality bidirectional sequences; whereas *psbK–psbI* showed good discriminatory power, but had the lowest sequencing success in these trials, and substantial problems generating bidirectional reads.

Choosing a plant barcode from the 3 remaining candidate loci was more difficult. Individually, *trnH–psbA*, *rbcL*, and *matK* possess attributes that are highly desirable in a plant DNA barcoding system, although none of the 3 loci fits all 3 criteria perfectly. As reported elsewhere (7), *trnH–psbA* demonstrated good amplification across land plants with a single pair of primers (93% for angiosperms; Fig. 1*A*) and high levels of species discrimination. However, problems obtaining high quality bidirectional sequences are the primary limitation for this locus. In addition, *trnH–psbA* has a median length of 418 bp (IQR = 296–500 bp) in the dataset examined here, which is well-suited for DNA barcoding, but its upper length of >1,000 bp in some monocot (3) and conifer (11) species can lead to problems obtaining bidirectional sequences without using taxon-specific internal sequencing primers.

Among plastid regions, *rbcL* is the best characterized gene. Improvements in primer design make it easily retrievable across land plants (8) and it is well suited for recovery of high-quality bidirectional sequences. Although not the most variable region (Fig. 1*B*), it is a frequent component of the best performing multi-locus combinations for species discrimination (Fig. 1*C*).

matK is one of the most rapidly evolving plastid coding regions and it consistently showed high levels of discrimination among angiosperm species (Fig. 1*C*) (8, 9). Mixed reports have been published regarding the universality of *matK* primers, ranging from routine success (9) to more patchy recovery (7, 8), which has led to reservations about this locus by some researchers. In the current study, 90% of the angiosperm samples tested were successfully amplified and sequenced using a single primer pair (Fig. 1*A*). Success in gymnosperms (83%) and particularly cryptogams (10%) was more limited, even when multiple primer sets were used.

In summary, *rbcL* offers high universality and good, but not outstanding discriminating power, whereas *matK* and *trnH–psbA* offer higher resolution, but each requires further development work. Primer universality needs improvement for *matK* in some clades, and *trnH–psbA* does not consistently provide bidirectional unambiguous sequences, often requiring manual editing of sequence traces. Thus, no single locus meets CBOL's data standards and guidelines for locus selection, and as a result a synergistic combination of loci is required.

One option preferred by some researchers in the CBOL Plant Working Group was a 3-locus barcode of *matK*+*rbcL*+*trnH-psbA*, to allow further testing of these loci. Based on the relative performance of the 3 loci, the best 2-locus barcode could be selected at a later date. The majority preference, however, was to select a 2-locus barcode to (a) avoid the increased costs of sequencing 3 loci rather than 2 in very large sample sets, and (b) prevent further

Downloaded from https://www.pnas.org by 205.139.151.11 on August 6, 2025 from IP address 205.139.151.11.

Tecan Ex. 2116, Page 2 of 4
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

delays in implementing a standard barcode for land plants. In the datasets examined here, sequencing 3 loci did not improve discrimination beyond the best performing 2-locus barcodes.

Among the 2-locus barcode combinations, *rbcL*+*matK* was the majority choice for several reasons. High-quality sequences of *rbcL* are easily retrievable across phylogenetically divergent lineages, and it performs well in discrimination tests in combination with other loci. Developing amplification strategies for *matK* was considered an investment with better prospects for return than solving the problem of sequence quality in *trnH–psbA* caused by mononucleotide repeats (13). Recent primer development for *matK* has improved its recovery from angiosperms, and so prospects for further improvement in angiosperms and other land plant groups seem reasonable, analogous to the extensive improvements made to primer sets for *CO1* for animal DNA barcoding (14).

We therefore propose *rbcL*+*matK* as the standard barcode for land plants. This combination represents a pragmatic solution to a complex trade-off between universality, sequence quality, discrimination, and cost. Using *rbcL*+*matK* in the sample set examined here, species discrimination was successful in 72% of cases, with the remaining species being matched to groups of congeneric species with 100% success. Given the logistical difficulties of undertaking identifications with some ≈400,000 species of land plant, this 2-locus barcode offers the opportunity to harness high-throughput automated sequencing technologies to establish a powerful universal framework for DNA-based identification of plants.

The unique identification to species level in 72% of cases and to 'species groups' in the remainder will be useful for many applications of DNA barcoding such as studies of plant-animal interactions (15), establishing whether plant products in international trade belong to protected species (9, 16, 17), discriminating among seedlings to establish forest regeneration dynamics, or undertaking large-scale biodiversity surveys with limited access to taxonomic expertise. A particular strength of the barcoding approach is that these identifications can be made with small amounts of tissue from sterile, juvenile or fragmentary materials from which morphological identifications are difficult or impossible (18). In addition, it is important to emphasize that the discriminatory power of this standard barcode will be higher in situations that involve geographically restricted sample sets, such as studies focusing on the plant biodiversity of a given region or local area (19, 20).

A future challenge for DNA barcoding in plants is to increase the proportion of cases in which unique species identifications are achieved. In the short term, where further resolution and universality are required, we envisage that the core *rbcL*+*matK* barcode will be augmented in individual projects from a flexible short-list of supplementary loci including the noncoding plastid regions examined here (*trnH–psbA*, *atpF–atpH*, and *psbK–psbI*), and the *trnL* intron which has been advocated for situations involving highly degraded tissue (19). The rapidly evolving internal transcribed spacers of nuclear ribosomal DNA also represent a useful supplementary barcode in taxonomic groups in which direct sequencing of this locus is possible (21). Moving beyond these currently available supplementary barcodes, ongoing advances in sequencing technologies and the concomitant accumulation of genomic and transcriptomic sequence data from plants will greatly increase opportunities for targeting the nuclear genome as a source of informative characters.

There is little doubt that the approaches used in plant DNA barcoding will be refined in future (22). However, the key foundation step for plant barcoding is in reaching agreement on a standard set of loci to enable large-scale sequencing and the development of a global plant barcoding infrastructure. The broad community agreement presented here, to sequence *rbcL* and *matK* as a standard 2-locus barcode, is thus an important step in establishing a centralized plant barcode database as a tool for taxonomy, conservation, and the multitude of other applications (23) that require identification of plant material.

Materials and Methods

Plant Materials. We used a total of 907 samples from 550 species representing the major lineages of land plants (including 670/445 angiosperm, 81/38 gymnosperm, and 156/67 cryptogam samples/species) to evaluate the candidate barcoding loci (Fig. S1, Fig. S2, and Table S1; cryptogams are defined here as all non–seed bearing embryophytes).

Universality. To provide directly comparable information on universality and trace quality (see below), we generated de novo sequence data from 190 samples (including 170 angiosperms) at the Canadian Centre for DNA Barcoding (CCDB), University of Guelph, using a single primer pair per locus (Table S1). We used this dataset to quantify universality in angiosperms. As amplification and sequencing success is typically lower in nonangiosperm land plants, which often require different primer sets, we compiled existing data on amplification and sequencing success from different laboratories as an indicator of success for these groups ($n = 81$ for gymnosperms; $n = 156$ for cryptogams; Table S1). Our assessments of universality simply record whether sequence data were obtained, regardless of the amount of manual trace editing required or the extent of read bidirectionality. Full details of molecular methods are available from the corresponding author on request.

Sequence Quality and Coverage. To assess suitability for bidirectional sequencing with minimal requirement for manual editing of sequences, we examined the quality of the de novo generated sequence traces via the CCDB automated informatics pipeline. Using a window size of 20 bp, segments with >2 bp showing <20 QV were trimmed. The amount of high-quality sequence data recovered was defined such that *both* the forward and reverse reads should have a minimum length of 100 bp, a minimum average QV of 30, and the post-trim lengths should be >50% of the original read length; the assembled contig should have >50% overlap in the alignment of the forward and reverse reads with <1% low-quality bases (<20QV) and <1% internal gaps and substitutions when aligning the forward and reverse reads. These quality control criteria were selected as a pragmatic set of thresholds to discriminate higher quality sequences from lower quality sequences. Various permutations of the parameters resulted in the same general conclusions (*rbcL*, *rpoC1*, and *rpoB* performed well, *matK* was intermediate, and fewer high-quality bidirectional sequences were obtained from *trnH–psbA*, *psbK–psbI*, and *atpF–atpH*).

Discrimination. To evaluate species discrimination we focused on samples from which all 7 loci were successfully sequenced (397 samples, all seed plants). We restricted assessment of discrimination success to species where multiple individuals were sampled from multiple congeneric species (259 samples of 95 species from 34 genera). Although not counted in the discrimination success statistics, a further 104 singleton-sampled species congeneric with the above, and 34 singleton-sampled species from 21 other genera were included to serve as potential sources of discrimination failure. Using the same samples for all 7 loci allowed us to directly compare the *relative* discriminatory power of the different loci. We considered discrimination as successful if the minimum uncorrected interspecific p-distance involving a species was larger than its maximum intraspecific distance [all distances were calculated from pairwise global alignments counting unambiguous base substitutions only (24)]. We evaluated species discrimination for multiple loci by summing the components of the distance measure for all possible 2–7 locus combinations and recording the success of each multi-locus combination. We used the binomial distribution to calculate 95% confidence intervals to establish whether performance differences between loci and locus combinations were statistically significant. Species discrimination assessments were then repeated on a dataset of 907 individuals/550 species that included samples successfully sequenced for some, but not all loci. Multi-locus combinations were not evaluated in this dataset because of large numbers of zero-distances introduced by individuals being represented by mutually exclusive loci.

ACKNOWLEDGMENTS. We thank David Schindel for comments, Sergey Sheetlin for data formatting, and George Weiblen for plant material. This work was supported by the Alfred P. Sloan Foundation, Gordon and Betty Moore Foundation, Genome Canada, Scottish Government's Rural and Environment Research and Analysis Directorate, Royal Society, South African National Research Foundation, Intramural Research Program of the National Library of Medicine, National Institutes of Health, and Consortium for the Barcode of Life.

Downloaded from https://www.pnas.org by 205.139.151.11 on August 6, 2025 from IP address 205.139.151.11.

Tecan Ex. 2116, Page 3 of 4
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029

SEE COMMENTARY

1. Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proc R Soc Biol Sci SerB* 270:313–321.
2. Kress JW, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proc Natl Acad Sci USA* 102:8369–8374.
3. Chase MW, et al. (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon* 56:295–299.
4. Ford CS, et al. (2009) Selection of candidate DNA barcoding regions for use on land plants. *Bot J Linn Soc* 159:1–11.
5. Pennisi E (2007) Taxonomy. Wanted: A barcode for plants. *Science* 318:190–191.
6. Ledford H (2008) Botanical identities: DNA barcoding for plants comes a step closer. *Nature* 451:616.
7. Kress WJ, Erickson DL (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.
8. Fazekas AJ, et al. (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* 3:e2802.
9. Lahaye R, et al. (2008) DNA barcoding the floras of biodiversity hotspots. *Proc Natl Acad Sci USA* 105:2923–2928.
10. Erickson DL, Spouge J, Resch A, Weigt LA, Kress JW (2008) DNA barcoding in land plants: Developing standards to quantify and maximize success. *Taxon* 57:1304–1316.
11. Hollingsworth ML, et al. (2009) Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Mol Ecol Res* 9:439–457.
12. Seberg O, Petersen G (2009) How many loci does it take to DNA barcode a crocus? *PLoS ONE* 4:e4598.
13. Devey DS, Chase MW, Clarkson JJ (2009) A stuttering start to plant DNA barcoding: Microsatellites present a previously overlooked problem in non-coding plastid regions. *Taxon* 58:7–15.
14. Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN (2007) Universal primer cocktails for fish DNA barcoding. *Mol Ecol Notes* 7:544–548.
15. Jurado-Rivera JA, Vogler AP, Reid CAM, Petitpierre E, Gómez-Zurita J (2009) DNA barcoding insect-host plant associations. *Proc R Soc Biol Sci SerB* 276:639–648.
16. Ogden R, et al. (2008) SNP-based method for the genetic identification of ramin *Gonystylus* spp. timber and products: Applied research meeting CITES enforcement needs. *Endang Species Res* doi 10.3354/esr00141.
17. Little DP, Stevenson DW (2007) A comparison of algorithms for the identification of specimens using DNA barcodes: Examples from gymnosperms. *Cladistics* 23:1–21.
18. Valentini A, Pompanon F, Taberlet P (2008) DNA barcoding for ecologists. *Trends Ecol Evol* 24:110–117.
19. Taberlet P, et al. (2006) Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. *Nucleic Acids Res* 35:e14.
20. Janzen DH (2005) in *Plant conservation*: *A natural history approach*, eds Krupnick G, Kress WJ (University of Chicago Press, Chicago), pp ix–xiii.
21. Feliner GN, Rosselló JA (2007) Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol Phylogenet Evol* 44:911–919.
22. Fazekas AJ, et al. (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Mol Ecol Res 9 S* 1:130–139.
23. Newmaster SG, Ragupathy S, Janovec J (2009) A botanical renaissance: State-of-the-art DNA bar coding facilitates an Automated Identification Technology system for plants. *Int J Comp Appl Tech* 35:50–60.
24. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 19:1792–1797.

ECOLOGY

Downloaded from https://www.pnas.org by 205.139.151.11 on August 6, 2025 from IP address 205.139.151.11.

Tecan Ex. 2116, Page 4 of 4
QIAGEN Sciences, LLC v. Tecan Genomics - IPR2025-00029