LANKESTERIANA 7(1-2): 200-203. 2007.

FINDING A SUITABLE DNA BARCODE FOR MESOAMERICAN ORCHIDS

GUILLAUME GIGOT[1,3], JONATHAN VAN ALPHEN-STAHL[1], DIEGO BOGARIN[2], JORGE WARNER[2], MARK W. CHASE[1] & VINCENT SAVOLAINEN[1]

[1]Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, TW9 3DS, United Kingdom.
[2]Jardín Botánico Lankester, Universidad de Costa Rica. P.O. Box 1031-7050 Cartago, Costa Rica, A.C.
[3]Author for correspondence: g.gigot@kew.org

KEY WORDS: DNA barcoding, orchids, Costa Rica, plastid genome, coding region, *trnH-psbA*

Introduction

Recently, DNA barcoding has emerged as an effective tool for species identification. This has the potential for many useful applications in conservation, such as biodiversity inventories, forensics and trade surveillance. It is being developed as an inexpensive and rapid molecular technique using short and standardized DNA sequences for species identification. The core idea of DNA barcoding is based on the fact that short pieces of DNA can be found that vary only to a minor degree within species, such that this variation is much less than between species (Savolainen *et al*. 2005). As proposed by Hebert *et al.* (2003), the DNA barcoding system for animals has been based upon sequence diversity in mitochondrial cytochrome *c* oxidase subunit 1 (COI or *cox1*). However, in land plants, the *cox1* gene has too low a rate of DNA sequence change to be used for species-level discrimination. The plastid genome of plants seems to be a better candidate for DNA barcoding, with enough variation to distinguish species and at the same time less intra- than inter-specific variability (Chase 2005, Cowan 2006). In 2005, Kress *et al.* proposed a non-coding plastid region, the *trnH-psbA* spacer, as a good barcode candidate. The Consortium for the Barcoding of Life (CBOL), via the Plant Working Group, has established another strategy to find a universal DNA barcode for land plants. A subset of six coding regions has been selected and is currently being tested in various plant taxa.

Our study is part of a project funded by the Darwin Initiative for the Survival of Species, which promotes biodiversity conservation and sustainable use of resources around the world (http://www.darwin.gov.uk). This project, based on a partnership between several academic and governmental authorities in Costa Rica with the Royal Botanic Gardens, Kew, in the UK, aims to record orchid diversity, establish long-term monitoring sites and undertake a pilot study on DNA barcoding for conservation and trade surveillance. Although some approaches to identify a DNA barcoding approach for land plants focused on a wide range of species around the world, e.g. the work lead by the Plant Working Group of CBOL, our work concentrates on a limited geographical area, Costa Rica, and a hyper-diverse family of plants, orchids. Costa Rica has one of the richest orchid floras in the world, with over 1300 species on a relatively small territory of 51,000 km^2. In spite of the fact that this country has a well-developed network of protected areas, with over 25% of its territory composed of protected forests and reserves, orchid floras remain under constant threat from factors such as deforestation and illegal trade. Furthermore, orchids are well known to be difficult to identify, particularly when they are sterile. Therefore, the use of a rapid and standardized DNA-based identification tool will be invaluable for many applications in conservation and to enforce international conventions such as the Biodiversity Convention (CBD) and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). Hence, among other activities of this project, we are currently working on the development of a DNA barcode for Mesoamerican orchids, in particular Costa Rican species.

Nuclear regions such as the internal transcribed spacer of the ribosomal DNA (ITS), although often highly variable in angiosperms, are not a practical option in several groups and show certain functional

limitation for DNA barcoding (Kress 2005). Both coding and non-coding plastid regions present various advantages (monomorphy, high copy number and highly diagnostic) and appear to be good candidates. One aim of barcoding is to find a "barcoding gap" between the intra- and inter- specific variation for the proposed regions (Meyer & Paulay 2005). We present here our preliminary results based on the comparison of different DNA plastid regions: the non-coding *trnH-psbA* spacer and five coding regions.

Materials and methods

We used standardized protocols for the PCR amplification and the sequencing available online on the RBG Kew website (http://www.kew.org/barcoding/protocols.html); all DNA samples came from the Kew DNA Bank (http://www.rbgkew.org.uk/data/dnaBank/homepage.html). The sampling covers 74 taxa representing 50 Mesoamerican orchid species and three temperate species as outgroups (from the North Temperate genus *Dactylorhiza*, Orchideae, Orchidoideae). We selected 47 species Costa Rican species and three species from other countries with a more northern distribution in Mesoamerica (from Mexico to Nicaragua).

To evaluate intra-specific variability, eleven of these species had multiple accessions (from two to seven). From the plastid genome, we sequenced the non-coding region *trnH-psbA* and portions of five DNA coding regions that have been put forward by the Plant Working Group of CBOL as potential universal barcodes for land plants, including *accD*, *rpoC1*, *rpoB*, *matK* and *ndhJ*. Altogether, these regions represent an aligned combined matrix of 3698 base pairs (bp) for 74 taxa.

We evaluated the inter- and intra-specific variation from a genetic distance matrix constructed using pair-wise Kimura 2 parameter (K2P) distances. The K2P model was used because it is simple and takes into account variable transition and transversion frequencies. Genetic distance between terminal taxa and their closest sister was used to characterize inter-specific divergence. The two most genetically distant individuals within each species were chosen to represent intra-specific divergence. We compared phylogenetic trees constructed using neighbour joining and parsimony methods. We also combined gene regions to evaluate the potential of a multi-locus barcode.

Results and discussion

Amplification was generally successful with all the regions tested. The only region that presented significant difficulties was *trnH-psbA;* there were alignment problems due to high levels of length variation. The sequence variability within and between species for all gene regions appears to overlap considerably, and, thus, these data do not show any evidence that there is a barcoding gap. Species groupings within neighbour joining and parsimony trees showed no topological differences. At the intra-specific level, the three gene regions that provided the greatest resolution were *matK*, *trnH-psbA* and *rpoB*, grouping over 50% of the eleven species with replicates into monophyletic groups. Among all combinations of regions tested as multi-locus barcodes, a "triplet" of *rpoC1*, *rpoB* and *matK* appeared to provide the best result and grouped all accessions of individuals correctly (Table 1).

Conclusion

As has been found in many plant groups (palms etc.), orchids exhibit low inter-specific sequence divergence, and there is no "barcode gap" between intra- and inter-specific data. However, results from the regions evaluated here show it is possible to

TABLE 1. Number of intra-specific species groupings per gene region from a neighbour joining tree (based on 11 species with replicates).

All regions are coding except for *trnH-psbA*.

Gene regions	Number of species groupings
accD	3 (27.3%)
matK	10 (90.9%)
ndhJ	1 (9.1%)
rpoB	6 (54.5%)
rpoC1	4 (36.4%)
trnH-psbA	8 (72.7%)
Triplet 1 (*rpoC1+rpoB+matK*)	11 (100%)
Triplet 2 (*rpoC1+matK+trnH-psbA*)	10 (90.9%)
Triplet 3 (*rpoB+matK+trnH-psbA*)	10 (90.9%)

group species replicates together, which is a basic requirement for a barcode identification tool. From the NJ reconstruction, the three best regions presenting the highest sequence variation and the best resolution at the species level are *rpoB*, *trnH-psbA* and *matK*.

It is clear that no single region will be sufficient as an efficient and universal barcode for orchids. A multi-locus barcode, based on two or three plastid regions, seems to be the most realistic and effective solution for the identification of Mesoamerican orchids. Our results show that a "triplet" of regions would be successful with a combination of regions like *rpoB*, *matK*, *trnH-psbA* or *rpoC1*. The next step for a multi-barcode will depend on the choice of using only coding regions or including a non-coding gene like *trnH-psbA*, although this gene presents practical complications with alignment.

ACKNOWLEDGMENTS. We thank the Darwin Initiative, the Consortium for the Barcode of Life, the Alfred P. Sloan and Gordon and Betty Moore Foundations.

LITERATURE CITED

Chase, M.W., N. Salamin, M. Wilkinson, J. M. Dunwell, R.P. Kesanakurthi, N. Haidar & V. Savolainen. 2005. Land plants and DNA barcodes: short-term and long-term goals. Phil. Trans. R. Soc. B 360: 1889–1895.

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

Hebert, P.D.N., S. Ratnasingham & J.R. De Waard. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. B 270: S96–S99.

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

Meyer, C.P. & G. Paulay. 2005. DNA barcoding: Error rates based on comprehensive sampling. PLoS Biol 3: 2229-2238.

Savolainen, V., R.S.Cowan, A.P. Vogler, G.K.Roderick & R. Lane. 2005 Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Phil. Trans. R. Soc. B 360: 1805–1811.

Guillaume Gigot was first educated at the University of Montpellier (France) where he studied evolution and ecology. He was then awarded his diploma of engineering in agronomy and environment at "Grande Ecole" in Paris. After working on several research projects in tropical ecology and population genetics in France, he started at the Royal Botanic Gardens, Kew in 2005 as Darwin Initiative Project Officer. He is currently in charge of the coordination and management of a project regarding orchid biodiversity and DNA barcoding in collaboration with the Lankester Botanical Garden in Costa Rica.

Jonathan van Alphen-Stahl completed his BSc in Biology, Earth and Environmental Sciences at the University of Cape Town in 2001. He went on to do his honours in Botany and then completed his Masters degree in Systematics and Biodiversity Science at the University of Cape Town, with his dissertation involving the phylogenetics and phylogeography of the Helmeted Guinea fowl. He worked as biodiversity officer at the University of Pretoria before moving to the Royal Botanic Gardens, Kew. He is currently working as a data analyst on DNA Barcoding at Kew.

Diego Bogarín obtained his degree in Biology at the University of Costa Rica. He is a researcher at Lankester Botanical Garden interested in taxonomy and systematics of neotropical Orchidaceae. Recently, he is developing floristic projects for conservation in Costa Rican National Protected Areas System. He started in 2005 as Darwin Initiative Project Implementation Officer in Costa Rica for the project "Conservation and monitoring of Meso-American orchids", in collaboration with Royal Botanic Gardens, Kew.

Mark Chase received his undergraduate degree from Albion College, Michigan and his Ph.D. was from the University of Michigan (Ann Arbor) in 1985. His thesis was a monograph of *Leochilus* (Orchidaceae). He carried out post-doctoral research in molecular biology with Jeff Palmer at the University of Michigan. He then moved to the University of North Carolina and then after four year to the Royal Botanic Gardens, Kew, where he set up the program in molecular systematics. He became a member of the Royal Society in 2003 and Keeper (Director) of the Jodrell Laboratory in 2006.

Jorge Warner obtanied his master degree at the University of Costa Rica. Actually, he is the Director of Lankester Botanical Garden and currently leader of the project "Conservation and monitoring of Meso-American orchids", sponsored by Darwin Initiative and developed in collaboration with Royal Botanic Gardens, Kew. His main interests are the biology and conservation of Costa Rican epiphytes.

Vincent Savolainen is a Plant Molecular Systematist and Deputy/Acting-Head of the Molecular Sytematics Section, at the Jodrell Laboratory, Royal Botanic Gardens, Kew. He received his PhD in Biology in 1995, from the University of Geneva, Switzerland, specializing in Molecular Phylogenetics and Evolution. His research interests include angiosperm phylogeny, Tree of Life, evolutionary processes and phylogenetics patterns, speciation and DNA Barcoding. He is currently leader of two projects funded by the Darwin Initiative for the Survival of Species in Southern Africa and Central America and several other European projects. He is presently a Fellow of the Linnean Society of London and was awarded the Bicentenary Medal of the Society in 2006.