Ciencia y Tecnología, 27(1 y 2): 24-34 ,2011
ISSN: 0378-0524

EVALUATION OF THREE CHROROPLASTIC MARKERS FOR BARCODING AND FOR PHYLOGENETIC RECONSTRUCTION PURPOSES IN NATIVE PLANTS OF COSTA RICA

Milton Vindas-Rodríguez[1], *Keilor Rojas-Jiménez*[2,3*], *Giselle Tamayo-Castillo* [2,4].

[1]Escuela de Biología, Instituto Tecnológico de Costa Rica; [2]Instituto Nacional de Biodiversidad, Costa Rica [3]Biotec Soluciones Costa Rica, S.A. [4]Escuela de Química, Universidad de Costa Rica.

Recibido 6 de diciembre, 2010; aceptado 30 de junio, 2011

Abstract

DNA barcoding has been proposed as a practical and standardized tool for species identification. However, the determination of the appropriate marker DNA regions is still a major challenge. In this study, we extracted DNA from 27 plant species belonging to 27 different families native of Costa Rica, amplified and sequenced the plastid genes *matK* and *rpoC1* and the intergenic spacer *trnH-psbA*. Bioinformatic analyses were performed with the aim of determining the utility of these markers as possible barcodes to discriminate among species and for phylogenetic reconstruction. From the markers selected, the *trnH-psbA* spacer was the most variable in terms of genetic distance and the most promising region for barcoding. However, it presented a limited use for constructing phylogenies due to the complexity of its alignment. The locus *matK* was less variable but was also useful for species discrimination and for phylogenetic tree generation. The *rpoC1* region was highly conserved and suitable for phylogenetic studies, but presented a limited utility as a barcode. The marker combination *matK* and *rpoC1* provided the best resolution for establishing valid phylogenetic relationships among the analyzed plant families. In conclusion, more than one marker should be used for plant barcoding purposes, to provide complementary and variable information for the interespecific species discrimination.

Resumen

Se ha propuesto el empleo de un código de barras de ADN como una herramienta práctica y estandarizada para la identificación de especies. Sin embargo, la determinación de los marcadores moleculares apropiados se constituye en todo un reto. En este estudio se ha extraído el ADN de 27 especies de plantas que pertenecen a 27 familias distintas, nativas de Costa Rica, de las cuales se amplificaron y secuenciaron los marcadores *matK* y *rpoC1* del genoma del plastidio y el espaciador intergénico *trnH-psbA*. Se realizaron análisis bioinformáticos con el propósito de determinar la utilidad de estos marcadores como posibles códigos de barra para discriminar entre especies y en la reconstrucción filogenética. De los marcadores seleccionados, el espaciador *trnH-psbA* fue el más variable en términos de distancia genética y la región más prometedora como código de barras. Sin embargo, presenta limitaciones en la construcción de filogenias debido a la complejidad del alineamiento. El locus *matK* fue menos variable, pero más útil en la discriminación de especies y en la generación de árboles filogenéticos. La región *rpoC1* fue altamente conservada y útil para estudios filogenéticos, pero de utilidad limitada para ser empleada como código de barras. La combinación de los marcadores *matK* y *rpoC1* provee la mejor resolución para establecer relaciones filogenéticas dentro de las familias de plantas analizadas. En conclusión, se requiere más de un marcador molecular para aplicaciones de

* Corresponding author: keilorrojas@yahoo.com

código de barras y para proveer información complementaria variable para la discriminación de especies inter-específicas.

Key words: DNA barcoding, molecular phylogeny in plants, *mat*K, *rpo*C1, *trn*H-*psb*A.

Palabras clave: Código de barras de ADN, filogenia molecular en plantas, *matK*, *rpoC1*, ***trnH-psbA***

I INTRODUCTION

The Costa Rican strategy for biodiversity conservation rests on three pillars: save a portion of the country in at least 25% of the territory, promote the generation of knowledge to understand and improve the management of those resources and finally, use them in a sustainable manner [1]. This approach is of key importance in a world of increasing population and in which the loss of organisms caused by human impact is threatening the natural balance. One important aspect of this strategy is the opportune identification of species to increase our understanding of the delicate equilibrium that prevails within their ecosystems. However, the limitations inherent of classifications and identification systems and the decline of taxonomists suggest the need for complementary approaches to modernize taxonomy. To overcome this situation, methods for species identification based on analysis of DNA sequences have been developed, which have taken a qualitative leap in terms of speed and accuracy to identify species [2].

The genetic barcoding is based on nucleotidic variability within standardized and short regions of genome. This allows the discovery and identification of species and also can be used for measuring diversity, ecological studies, and forensic analysis [3, 4]. To be functional and universal, the barcode must meet the following criteria: Firstly, it must exhibit high interspecific but low intraspecific divergence. In addition, the marker must be standardized within the same DNA region in different taxonomic groups. Thirdly, it should contain enough information to easily assign the organism to a species level. It is also desirable for the marker to be a coding region, useful for studies of phylogenetic relationships and molecular evolution. Finally, the barcode must be extremely robust, with conserved flanking sites for developing universal primers so as to facilitate amplification and sequencing [3, 5, 6].

In animals, the *cox*1 gene has been widely used as a barcode because of its easy sequencing while being highly informative [7]. Among plants, particularly angiosperms, the selection of a standard barcode is more difficult mainly due to the evolutionary trends of each of the three plant genomes. The nuclear genome presents polyploidy that often results in multiple and heterogeneous copies of the markers. The mitochondrial DNA, compared with other eukaryotes, is subject to rearrangements, has a very low rate of substitution and is more likely to migrate to the nucleus [8]. Conversely, the chloroplast genome is haploid, uniparentally inherited, structurally stable, nonrecombining and has a high rate of substitution, which makes it suitable for tracking speciation events. Nevertheless, it has been determined that a system based on one barcode could succeed or not depending on the intraspecific variation of the taxonomic groups. Therefore, more than one barcode may be necessary to increase the power of discrimination between closely related species [9, 10].

In this study we evaluated the performance of three chloroplast markers for DNA barcoding purposes and also for complementary constructing phylogenetic relationships in plant families of Costa Rica. The markers selected were the *mat*K and *rpo*C1 genes and the intergenic spacer *trn*H-*psb*A. The *mat*K was selected for its easy amplification and alignment besides its desirable high interspecific but low intraspecific variation [6]. The *trn*H-*psb*A region was selected

because of its high variability, short size and because the published primers seem to amplify in almost all angiosperm taxa [6, 11]. The plastid region, *rpo*C1, was chosen for its good performance in terms of being amplified with a limited range of PCR conditions and primer groups and for the reliable discrimination achieved among species of different plant groups [11]. After collecting samples in the field, we extracted, amplified and sequenced DNA for this study. Several bioinformatics tools were used to evaluate the potential of these regions as barcodes and for phylogenetic studies; comparisons against the Genbank using BLAST algorithm were performed to establish the most related species to the analyzed sequence [12, 13]. Intra- and interspecific genetic divergences were assessed by using pairwise calculations. Phylogenetic relationships among species were determined with Distance, Bayesian and Maximun Likelihood methods in order the construct the most probable trees [14, 15].

II MATERIALS AND METHODS

Sample collection. A total of 27 representative species of 27 different families (approximately 8% of Costa Rican plant diversity) were collected in several geographical locations of Costa Rica under permit number R-CM-INBio-44-2008-OT of the National Ministry of Environment. The samples consisted of approximately 10 g. of plant tissues, mainly leaves, cut with ethanol-sterilized scissors and preserved in hermetically sealed packages until transported to the laboratory where they were stored at -20º C. A description of the taxonomy and collecting sites of the samples is shown in Table 1.

DNA manipulations. DNA was extracted from plant material, using the DNeasy Plant Mini Kit (Qiagen ®, Valencia, California, USA) following the manufacturer's instructions. The DNA obtained was used as a template for the amplification of plastid genes *matK* and *rpoC1* and the intergenic spacer *trnH-psbA* with primers described in table 2, according to the combination proposed by Kress *et al*. [3]. The PCR conditions for *mat*K amplification were: initial denaturation of 3 min at 94 ºC, followed by 35 cycles of 1 min. denaturation at 94 ºC, 45 sec. hybridization at 48 °C, 1 min. extension at 72 ºC , 7 min. final extension. The conditions for the amplification of *rpo*C1 consisted of an initial denaturation of 3 min at 94 ºC, followed by 35 cycles of 1 min. denaturation at 94 ºC, 45 sec. hybridization at 48 °C, 1 min. extension at 72 ºC, 7 min. final extension. For amplifying the intergenic spacer *trn*H-*psb*A the conditions were: initial denaturation of 3 min at 94 ºC, followed by 35 cycles of 1 min. denaturation at 94 ºC, 30 sec. hybridization at 51 °C, 1 min. extension at 72 ºC, 7 min. final extension. The amplicons were purified using the QIAquick PCR Purification Kit (Qiagen) according to the specifications given by the manufacturer. Sequencing was performed at the sequencing facility of the Dana Farber Cancer Institute at Harvard Medical School at Longwood, Boston, Massachusetts, using the same primers as for PCR.

Sequence analysis. Sequences were assembled and edited with Seqman Pro tool of the DNASTAR lasergen8.0 program (GenBank accession GQ429055-GQ429145). Subsequently, a BLAST was conducted to compare the sequences obtained with those reported in the GenBank [12] to identify the species through the partial sequences of the markers, as well as to quantify the new information provided by this study.

Phylogenetic analysis. Sequences of each marker were aligned using multiple alignment programs: CLUSTAL W2 [16], Kalign [17] and MUSCLE [18] and then optimized with BIOEDIT [19]. The molecular phylogeny was constructed by using neighbor joining (NJ) distance methods

with the MEGA 4 Software [20]. The analysis of nucleotide frequencies and transition/transversion substitutions confirmed Tamura-Nei (TN93) as an appropriate distance correction model for the dataset. The statistical support for the phylogenetic branches was calculated with 1,000 bootstrap replications [21, 22]. In addition, Bayesian estimations were made to each proposed region as barcode using MrBayes 3.0b.4 program, establishing 1,000,000 generations, with a frequency sampling and printing between 100 and 4 chains for Monte Carlo Markov Chain analysis, which generated 10,000 trees. The stationary state was obtained about 8,000 trees, so a *burning* of 8,000 was selected; in this way, only 2,000 trees would be used to determine the consensus tree and the probabilities of each tree. The consensus trees were observed using TreeView [23]. Finally, maximum likelihood analysis were conducted to the alignments of each marker using the PAUP*4.0 [24] program to obtain consensus trees with a bootstrap support of 1,000, and the most appropriate model for database was selected with MODELTEST 3.7 [25].

III RESULTS

Sequence analysis. Results of the comparison of the three markers evaluated for each of the 27 species against the Genbank showed that 80% of the *matK* sequences presented corresponding sequences in the database, while it was not possible to find matches for the remaining 20% and for most of the sequences of *rpo*C1 and *trn*H-*psb*A. When analyzing the alignments statistics for each region and for their combinations, results showed that the marker that exhibited the greatest interspecific divergence was *trn*H-*psb*A with 1.02 ± 0.07 (Table 3). In addition, it was shown that sequences of *trn*H-*psb*A were unique-sized and presented the highest number of parsimony-informative sites (95.9%), which is quite relevant for rapid identification purposes. However, this region presented numerous indels that complex the sequence alignment and hence the phylogenetic reconstruction. The marker *mat*K presented an average distance of 0.29 ± 0.03 and 52.1% of parsimony-informative sites. This marker provided enough variation to distinguish to the species level when sequences were compared to their respective matches in the GenBank and also was useful for constructing reliable phylogenies. The *rpo*C1 marker was the most conserved presenting a low interspecific divergence of 0.13 ± 0.01 and 28.0% of parsimony-informative sites. The combination of *mat*K and *rpo*C1 produced an average distance of 0.17 ± 0.01 and a percentage of parsimony-informative sites of 41.2%, which lies in between the *mat*K and *rpo*C1 values. The combination of the three markers did not add significant variation to the alignment (0.165 ± 0.001).

Phylogenetic analysis. Several phylogenetic trees were generated for the three markers alone and combined using Distance, Bayesian and Maximum Likelihood methods. Results showed that the phylogenetic tree more consistent with the evolution of *Viridiplantae* proposed by the Angiosperm Phylogeny Group (APGII) classification system and additionally supported by the bootstrap was obtained with the combination of *rpo*C1 and *mat*K markers, using the neighbor joining method and applying a Tamura-Nei distance-correction (Figure 1). The Bayesian and Maximum Likelihood generated trees based on the combination of these markers also produced similar results.

The phylogenetic tree obtained clearly separated the two major groups; angiosperms and gymnosperms/lycopods. Furthermore, it was possible to distinguish the main monophyletic groups within angiosperms: Monocots, Magnolidae and the Eudicots. Within the clade of Monocots, *Carludovica drudei* and *Bomarea chiriquina*, which belong to Liliales and Pandanales, respectively, were placed as sisters while *Costus pulverulentus* and *Calathea lutea* were settled in nearby clades. *Dieffenbachia* sp. (Alimatales) was positioned in the lowest groups of this cluster. Within Magnolidae, it was possible to distinguish between Laurales represented by *Ocotea*

fulvescens and *Mollinedia pinchotiana* and Magnoliales represented by *Xylopia bocatorena*. The clade of Piperales (*Piper umbellatum*) was separated from these two orders with a lower bootstrap.

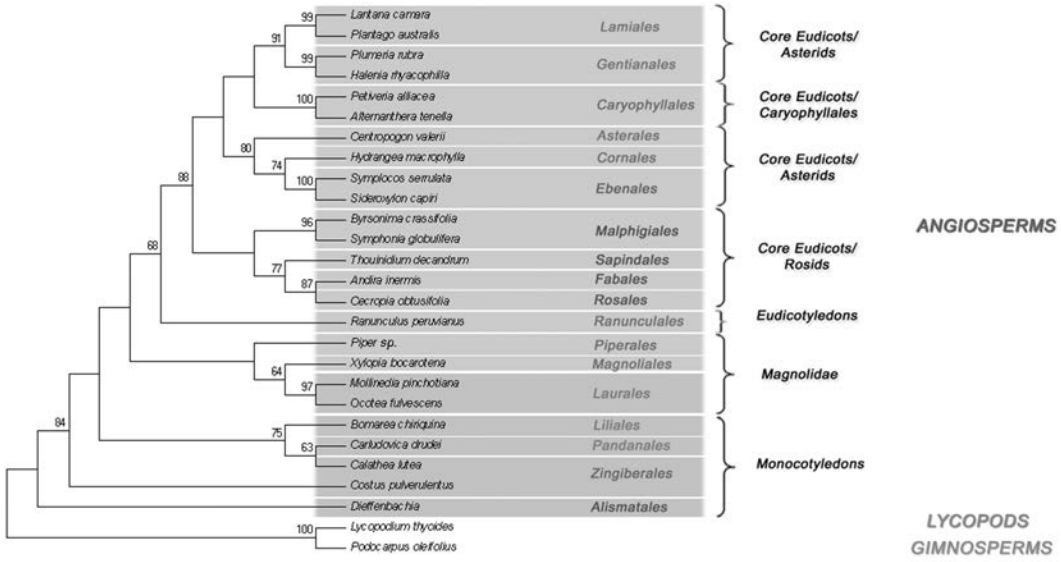

FIGURE 1. Phylogenetic relationships of the sampled plant species, determined using the neighbor joining method and applying a Tamura-Nei distance-correction. Percentage bootstrap support (1,000 replicates) is indicated at branching points. The order definition is based on the APGII [26].

Within the large group of dicotyledons angiosperms, two main clades were observed: the group of Eudicotyledons represented only by *Ranunculus peruvianus* and the Core Eudicots. The Core Eudicots was clearly subdivided into three subclades; Rosids, Caryophyllales and Asterids. At the Rosids clade, *Andira inermis* was placed close to *Cecropia obtusifolia* and *Thouinidium decandrum* and slightly separated to *Byrsonima crassifolia* and *Symphonia globulifera*. Within Caryophyllales, *Petiveria alliacea* and *Alternanthera tenella* were grouped as sisters with strong support. Finally, within Asterids the Lamiales clade represented by *Lantana camara* and *Plantago australis* were separated from Gentianales represented by *Plumeria rubra* and *Halenia rhyacophilla*. Within the other subclade of Asterales, Cornales represented by *Hydrangea macrophylla* were distinguished from Ebenales species *Symplocos serrulata* and *Sideroxylon capiri*.

The analysis of trees generated by Distance, Bayesian and Maximum likelihood methods, aligned with the sequences of *mat*K evidenced that it could not effectively separate angiosperms from other less evolved groups (Gymnosperms and Lycopods). Furthermore, the bootstrap values for all clades were smaller than those of the combination of *mat*K and *rpo*C1 (Table 4). In the case of the *rpo*C1 marker, similar trees were found using methods of distance, maximum likelihood and Bayesian. This region was highly conserved and clearly separated Angiosperms from Gymnosperms. The resolution of this marker at lower taxonomic levels was increased when combined with a more variable barcode such as *mat*K. However, the addition of a third marker (*trn*H-*psb*A) did not generate a significant increase in the resolution provided by the *mat*K and *rpo*C1 matrix.

IV DISCUSSION

Phylogenetic analysis. The combination of *rpo*C1 and *mat*K sequences exhibited the tree with the highest phylogenetic resolution, placing the 27 sampled species correctly, according to the system

proposed by the APGII. *Lycopodium thyoides* and *Podocarpus oleifolius* were separated from each other with a maximum resolution. The sequences from *rpo*C1 marker provided sufficient parsimony-informative sites to separate these groups less evolved from the large clade of angiosperms, resolution that was not evidenced with the *mat*K region. This confirmed previous results that mention *rpo*C1 as a marker that easily discriminate between angiosperms and non-angiosperms [11].

The sampled species within the group of the monocots *Costus pulverulentus* and *Calathea lutea* were placed closely because they belong to the order of Zingiberales while *Dieffenbachia* sp. was separated from this cluster for the reason that it belongs to Alismatales, a root order of monocots [27, 28]. Within Magnolids, *Mollinedia pinchotiana* and *Ocotea fulvescens*, which belong to the Lauraceae and Monimiaceae families, respectively, were properly located as sisters and clearly discriminated from *Xylopia bocarotena*, which belongs to Annonaceae family. This is consistent with previous studies that showed Magnoliales and Laurales as sister groups [29, 30]. Moreover, *Piper umbellatum*, which corresponds to Piperaceae, was distanced from Laurales and Magnoliales, coinciding with Zanis *et al*. [31], who suggested this order as brother of the distant group of Cannellales.

Within Eudicotyledons, *Ranunculus peruvianus* was separated from the rest of eudicots since Ranunculales is a basal order of eudicots and is hypothesized that it is monophyletic, according to the analysis of multiple DNA sequences such as *rbc*L, *atp*B and 18S rDNA [32, 33, 34]. Within the Core Eudicotyledons**,** particularly within the Rosids subgroup, *Andira inermis* of the order Fabales was close to *Cecropia obtusifolia* (Rosales order) while *Thouinidium decandrum* was distinguished of these last two orders, because it belongs to the Sapindales. This is consistent with the APGII [26] that indicates that Fabales and Rosales belong to the Eurosidas I subgroup, while the Sapindales belongs to Eurosidas II. In the Caryophyllales subgroup, *Alternanthera tenella* and *Petiveria alliacea*, were accurately classified as sisters in concordance with Cuénoud *et al*. [35] that established both species associated to the Core Caryophyllales. Within Asterids, *Symplocos serrulata* and *Sideroxylon capiri* were grouped as sisters, because both belong to the Ebenales order and were separated from *Hydrangea macrophylla* that belongs to Cornales, the root of all other Asterids [36]. *Lantana camara* and *Plantago australis* were placed as sisters, because both belong to the Lamiales order whereas *Plumeria rubra* and *Halenia rhyacophilla* were correctly grouped within Gentianales order. *Centropogon valerii*, classified in the Asterales order, particularly in another monophyletic clade known as Euasterids I, was separated from Euasterids II (Gentianales and Lamiales), coinciding with APGII [26].

Analysis of barcodes. This study suggests that the intergenic spacer *trn*H-*psb*A is a suitable option for plant DNA barcoding to discriminate specimens of different plant families since it was highly variable in terms of genetic distance and sequence length and because it showed an easy amplification for the variety of families involved. These results are consistent with other studies revealing this non-coding plastidic marker as a potential barcode for plants [9, 37, 38]. Conversely, the weakness of this marker consists in its limited potential for the phylogenetical reconstruction mainly due to its difficult alignment, as it was shown previously [39]. Nevertheless, an easy alignment is desirable, it is not required for a barcoding purposes since BLAST searches, for example, can directly associate the sequence of any species to the corresponding match in the Genbank. In this regard, a fundamental issue to address in order to convert the *trn*H-*psb*A region in a standard barcode for plants is increasing the amount of sequences in databases since currently only information of few species is available in the Genbank.

Although *mat*K marker showed a smaller divergence than *trn*H-*psb*A, accurate assignments were retrieved when comparing the sequences against the Genbank. This indicates that *mat*K provides enough information for interspecific differentiation, as confirmed in other studies [6]. In addition, it has been estimated that the pattern of variation of the second half of the 5′ region of the gene is particularly suitable for barcoding while the first half is useful for phylogenetic analyses in angiosperms [40]. This feature, suggests that this gene might be used as a dual barcode-phylogenetic marker. Nevertheless, the application of *mat*K as a standard plant barcode relies on the development of universal primers that effectively amplify this region across different groups of plants [11, 35]. The marker *rpo*C1 was highly conserved and offered adequate resolution to perform studies on the phylogenetic evolution of the species sampled. However, it presented the lowest sequence divergence, which limits its use for identification purposes [6]. In conclusion, complementary and variable information is required for an effective discrimination where a multiregion approach might provide additional parsimony-informative sites to identify all plant species improving the resolution of the DNA-barcoding [3, 10, 41].

V REFERENCES

[1] Tamayo, G.; Guevara, L.; Gámez, R. In: Bull, A. (Eds.), *Microbial Diversity and Bioprospecting*, Washington, D.C., American Society for Microbiology Press, 2004.

[2] Hebert, H.; Cywinska, A.; Ball, S.; deWaard, J., *Proc. R. Soc. London. B.* **2003**, *270*, 313-321.

[3] Kress, W.; Wurdack, K.; Zimmerm, E.; Weigt, L.; Janzen, D., *PNAS* **2005**, *102*, 8369-8374.

[4] Lanteri, A., *Rev. Soc. Entomol. Argent.* **2007**, *66*, 15-25.

[5] Taberlet, P.; Coissac, E.; Pompanon, F.; Gielly, L.; Miquel, C.; Valentini, A.; Vermat, T.; Corthier, G.; Brochmann, C.; Willerslev, E., *Nucleic Acids Res.* **2007**, *35*, 14-21.

[6] Lahaye R.; Van der bank, M.; Bogarin, D.; Warner, J.; Pupulin, F.; Gigot, G.; Maurin, O.; Duthoit, S.; Barraclough, T.; Savolainen, V., *PNAS* **2008**, *105*, 2923-2928.

[7] Hebert, P.; Gregory, T., *Syst. Biol.* **2005**, *54*, 852-859.

[8] Freudenstein, J.; Senyo, D.; Chase, M. In: Wilson, K. L.; Morrison, D. A. (Eds.), *Monocots: systematics and evolution*, Victoria, Australia, CSIRO, 2000.

[9] Shaw, J.; Lickey, E.; Beck, J.; Farmer, S.; Liu, W.; Miller, J.; Siripun, K.; Winder, C.; Schilling, E.; Small, R., *Am. J. Bot.* **2005**, *92*, 142-166.

[10] Newmaster, S.; Fazekas, R.; Steeves, A.; Janoverc, J., *Mol. Ecol. Notes* **2007**, *1*, 1-11.

[11] Chase, M., Cowan, R.; Hollingsworth, P.; Van den berg, C.; Madriñán, S.; Petersen, G.; Seberg, O.; Jorgsensen, T.; Cameron, K.; Carine, M.; Pedersen, N.; Hedderson, T.; Conrad, F.; Salazar, G.; Richardson, J.; Hollingsworth, M.; Barraclough, T.; Kelly, L.; Wilkinson, M., *Taxon* **2007**, *56*, 295-299.

[12] Altschul, S.; Madden, S.; Schaffer, A.; Zhang, J.; Miller, W.; Lipman, D., *Nucleic Acids Res.* **1997**, *25*, 3389-3402.

[13] Steinke, D.; Vences, M.; Salzburger, W.; Meyer, A., *Phil. Trans. R. Soc. B.* **2005**, *360*, 1975-1980.

[14] Saitou, N.; Nei, M., *Mol. Biol. Evol.* **1987**, *4*, 406-425.

[15] Lambert, D.; Baker, A.; Huynen, I.; Haddrath, O.; Hebert, P.; Millar, C., *J. Hered.* **2005**, *96*, 279-284.

[16] Larkin, M.; Blackshields, G.; Brown, N.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G., *Bioinformatics* **2007**, *23*, 2947-2948.

[17] Lassmann, T.; Sonnhammer, E., *BMC Bioinformatics* **2005**, *6*, 298.
[18] Edgar, R., *BMC Bioinformatics* **2004**, *5*, 113.
[19] Hall, T., *Nucl. Acids. Symp. Ser.* **1999**, *41*, 95-98.
[20] Tamura, K.; Dudley, J.; Nei, M.; Kumar, S., *Mol. Biol. Evol***.** **2007**, *24*, 1596-1599.
[21] Felsenstein, J., *Evolution* **1985**, *39*, 783-791.
[22] Farris, J.; Albert, V.; Kallersjo, M.; Lipscomb, D.; Kluge, A., *Cladistics* **1996**, *12*, 99-124.
[23] Page, R., *Comput. Appl. Biosci.* **1996**, *12*, 357-358.
[24] Swofford, D., *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4, Sinauer Associates, Sunderland, Massachusetts, 2003.
[25] Posada, D.; Crandall, K., *Bioinformatics* **1998**, *14*, 817-818.
[26] The Angiosperm Phylogeny Group, *Bot. J. Linean Soc.* **2003**, *141*, 399-436.
[27] Soltis, D. ; Soltis, P. ; Chase, M.; Mort, M.; Albach, D.; Zanis, M.; Savolainen, V.; Hahn, W.; Hoot, S.; Fay, M.; Axtell, M.; Swensen, S.; Prince, L.; Kress, W.; Nixon, K.; Farris, J., *Bot. J. Linean Soc.* **2000**, *133*, 381-461.
[28] Chase, M.; Soltis, D.; Soltis, P.; Rudall, P.; Fay, M.; Hahn, W.; Sullivan, S.; Joseph, J.; Molvray, M.; Kores, P.; Givnish, T.; Sytsma, K.; Pires, J. In: Wilson, K. L.; Morrison, D. A. (Eds.), *Systematics and evolution of monocots*, Proceedings of the 2nd International Monocot Symposium, Melbourne, CSIRO, 2000.
[29] Qiu, Y.; Lee, J.; Bernasconi-Quadroni, F.; Soltis, D.; Soltis, P.; Zanis, M.; Zimmer, E.; Chen, Z.; Savolainen, V.; Chase, M., *Nature* **1999**, *402*, 404-407.
[30] Qiu, Y., *Int. J. Plant Sci.* **2005**, *166*, 815-842.
[31] Zanis, M.; Soltis, D.; Soltis, P.; Qiu, Y.; Mathews, S.; Donoghue, M., *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6848-6853.
[32] Chase, M.; Soltis, M.; Olmstead, D.; Morgan, R.; Les, D.; Mishler, D.; Duvall, B.; Price, M.; Hills, H.; Qiu, Y.; Kron, K.; Rettig, J.; Conti, E.; Palmer, J.; Manhart, J.; Sytsma, K.; Michaels, H.; Kress, W.; Karol, K.; Clark, W.; Hedrén, M.; Gaut, B.; Jansen, R.; Kim, K.; Wimpee, C.; Smith, J.; Furnier, G.; Strauss, S.; Xiang, Q.; Plunkett, G.; Soltis, P.; Swensen, S.; Williams, S.; Gadek, P.; Quinn, C.; Eguiarte, L.; Golenberg, E.; Learn, G.; Graham, S.; Barrett, S.; Dayanandan, S.; Albert, V., *Ann. Missouri Bot. Gard.* **1993**, *80*, 528-580.
[33] Drinnan, A.; Crane, P.; Hoot, S., *Plant Syst. Evol.* **1994**, *8*, 93-122.
[34] Hoot, S.; Crane, P., *Plant Syst. Evol.* **1995**, *9*, 119-131.
[35] Cuénoud, P.; Savolainen, V.; Chatrou, L.; Powell, M.; Grayer, R.; Chase, M., *Am. J. Bot.* **2002**, *89*, 132-144.
[36] Bremer, B.; Bremer, K.; Heidari, N.; Erixon, P.; Anderberg, A.; Olmstead, R.; Källersjö, M.; Barkhordarian, E., *Mol. Phylogenet. Evol.* **2002**, *24*, 274–301.
[37] Chandler, G.; Bayer, R.; Crisp, M., *Am. J. Bot.* **2001**, *88*, 1675-1687.
[38] Miller, J.; Grimes, J.; Murphy, D.; Bayer, R.; Ladiges, P., *Syst. Botany* **2003**, *28*, 558-566.
[39] Hamilton, M.; Braverman, J.; Soria-Hernanz, D., *Mol. Biol. Evol.* **2003**, *20*, 1710-1721.
[40] Hilu, K.; Borsch, T.; Muller, K.; Soltis, D.; Soltis, P.; Savolainen, V.; Chase, M.; Powell, M.; Alice, L.; Evans, R.; Sauquet, H.; Neinhuis, C.; Slotta, T.; Rohwer, J.; Chatrou, L., *Am. J. Bot.* **2003**, *90*, 1758-1776.
[41] Cowan, R.; Chase, M.; Kress, W.; Savolainen, V., *Taxon* **2006**, *55*, 611-616.
[42] Sang, T.; Crawford, D.; Stuessy, T., *Am. J. Bot.* **1997**, *84*, 1120-1136.
[43] Tate, J.; Simpson, B., *Syst. Botany* **2003**, *28*, 723-737.

Acknowledgements

We would like to thank Luis Diego Vargas and Luis Guillermo Acosta from INBio for their collaboration during the design and execution of the sample collecting. We thank also Nelson Zamora for the initial selection of the material under study. This research was supported by the Korean Research Institute of Biosciences and Biotechnology (KRIBB) and by the NIH grant (U01 TW007404-01 ICBG). We thank Area de Conservación Guanacaste, Area de Conservación Pacífico Central, Area de Conservación Cordillera Volcánica Central and Area de Consevación Osa as well as the CONAGEBIO and the Ministry of Environment and Energy of Costa Rica for the collection permit (R-004-2004-OT-MINAE).

Appendix

TABLE 1
DESCRIPTION OF THE TAXONOMY AND RESPECTIVE COLLECTING SITES OF THE PLANT SAMPLES.

Family	Species	Conservation Area	Longitude	Latitude
Fabaceae/pap.	*Andira inermis*	Guanacaste	-85:40:47.7	10:51:28.7
Sapindaceae	*Thouinidium decandrum*	Guanacaste	-85:40:47.7	10:51:28.7
Sapotaceae	*Sideroxylon capiri*	Guanacaste	-85:40:47.7	10:51:28.7
Phytolaccaceae	*Petiveria alliacea*	Pacífico Central	-84:42:6.1	9:52:57.8
Campanulaceae	*Centropogon valerii*	Pacífico Central	-83:51:35.5	9:37:0.4
Ranunculaceae	*Ranunculus peruvianus*	Pacífico Central	-83:51:35.5	9:37:0.4
Podocarpaceae	*Podocarpus oleifolius*	Pacífico Central	-83:49:28.8	9:36:22.0
Lycopodiaceae	*Lycopodium thyoides*	Pacífico Central	-83:49:28.8	9:36:22.0
Plantaginaceae	*Plantago australis*	Pacífico Central	-83:49:28.8	9:36:22.0
Lauraceae	*Ocotea fulvescens*	Pacífico Central	-83:49:28.8	9:36:22.0
Marantaceae	*Calathea lutea*	Cordillera Volcánica Central	-83:51:24.0	10:11:52.0
Verbenaceae	*Lantana camara*	Cordillera Volcánica Central	-83:51:24.0	10:11:52.0
Cecropiaceae	*Cecropia obtusifolia*	Cordillera Volcánica Central	-83:51:24.0	10:11:52.0
Monimiaceae	*Mollinedia pinchotiana*	Cordillera Volcánica Central	-83:51:40.6	10:11:38.1
Clusiaceae	*Symphonia globulifera*	Cordillera Volcánica Central	-83:51:40.6	10:11:38.1
Cyclanthaceae	*Carludovica drudei*	Osa	-83:10:45.6	8:39:15.8
Apocynaceae	*Plumeria rubra*	Osa	-83:10:45.6	8:39:15.8
Malpighiaceae	*Byrsonima crassifolia*	Osa	-83:10:36.0	8:39:25.6
Costaceae	*Costus pulverulentus*	Osa	-83:10:36.0	8:39:25.6
Araceae	*Dieffenbachia sp.*	Osa	-83:10:36.0	8:39:25.6
Hydrangeaceae	*Hydrangea macrophylla*	Cordillera Volcánica Central	-84:14:0.5	10:10:13.4
Symplocaceae	*Symplocos serrulata*	Cordillera Volcánica Central	-84:14:0.5	10:10:13.4
Piperaceae	*Piper umbellatum*	Cordillera Volcánica Central	-83:59:12.0	10:3:30.0
Annonaceae	*Xylopia bocatorena*	Cordillera Volcánica Central	-83:51:49.2	10:11:7.2
Amaranthaceae	*Alternanthera tenella*	Pacífico Central	-84:35:23.3	9:48:11.4
Alstroemeriaceae	*Bomarea chiriquina*	Cordillera Volcánica Central	-83:50:14.3	9:58:56.4
Gentianaceae	*Halenia rhyacophila*	Pacífico Central	-83:49:20.8	9:36:21.8

TABLE 2
LIST OF THE PRIMERS USED IN THIS STUDY.

Genetic marker	Primer Sequence	Reference
***mat*K**		
390'f	CGATCTATTCATTCAATATTTC	[35]
1326'r	TCTAGCACACGAAAGTCGAAGT	[35]
***trn*H-*psb*A**		
psbA3'f	GTTATGCATGAACGTAATGCTC	[42]
trnH'r	CGCGCATGGTGGATTCACAATCC	[43]
***rpo*C1**		
LP1	TATGAAACCAGAATGGATGG	[11]
LP5	CAAGAAGCATATCTTGASTYGG	[11]

TABLE 3
STATISTICS OF THE ALIGNMENTS OF THE *LOCI* EVALUATED.

Primers	Alignment Length (bp)	Average percentage of nucleotide frequencies (T/C/A/G)	Average genetic distance of sequences	Parsimony-informative sites percentage
*mat*K	762	36.5 / 18.0 / 29.4 / 16.2	0.29(±0.03)	52.1
*rpo*C1	746	29.6 / 18.0 / 29.2 / 23,2	0.13 (±0.01)	28.0
*mat*K+*rpo*C1	1,573	33.0 / 18.0 / 29.4 / 19.6	0.17 (±0.01)	41.2
*trn*H-*psb*A	321	31.4 / 15.7 / 38.4 / 14.4	1.02 (±0.07)	95.9
*mat*K+*rpo*C1+*trn*H-*psb*A	1,776	32.8 / 17.7 / 30.4 / 19.1	0.165 (±0.01)	48.1

The alignments were generated with the program clustalw2 and the distance matrix estimated with MEGA4.

TABLE 4
BOOTSTRAP SUPPORT OF THE CLADES GENERATED FROM THE MATRICES OF THE MARKERS.

Clade	*matK*	*rpoC1*	*matK+rpoC1*	*matK+rpoC1 +trnH-psbA*
Angiosperms vrs Gimnosperms	N/A	100	100	100
Gimnosperms vrs Lycopodium	N/A	100	100	100
Magnolidae vrs Monocots	<50	<50	<50	90
Magnolidae vrs Eudicots	<50	<50	68	<50
Eudicots (Ranunculales) vrs Core Eudicots	N/A	<50	88	70
Rosids vrs Asterids /Caryophyllales	<50	<50	<50	91
Mollinedia pinchotiana vrs Ocotea fulvescens	<50	98	97	92
Laurales vrs Magnoliales	65	<50	64	68
Piperales vrs Magnoliales/Laurales	<50	<50	<50	N/A
Pandanales vrs Liliales	<50	<50	75	66
Rosales vrs Fabales	57	57	87	75
Alismatales vrs Zingiberales	<50	<50	84	100
Sapindales vrs Rosales	<50	<50	77	60
Malphigiales vrs Sapindales	<50	<50	<50	57
Byrsonima crassifolia vrs *Symphonia globulifera*	97	<50	96	95
Periveria alliacea vrs *Alternanthera tenella*	97	67	100	99
Plumeria rubra vrs *Halenia rhyacophilla*	83	89	99	99
Lantana camara vrs *Plantago australis*	99	90	99	99
Gentianales vrs Lamiales	54	63	91	81
Lamiales/Gentianales vrs Asterales/Cornales/Ebenales	<50	<50	<50	<50
Ebenales/Cornales vrs Asterales	<50	<50	80	79
Cornales vrs Ebenales	<50	<50	74	66
Symplocos serrulata vrs *Sideroxylon capiri*	86	94	100	99