

DNA Barcoding Bromeliaceae: Achievements and Pitfalls

Vitor Hugo Maia^{2,3}, Camila Souza da Mata^{1,3}, Luciana Ozório Franco¹, Mônica Aires Cardoso¹, Sérgio Ricardo Sodr  Cardoso¹, Adriana Silva Hemerly², Paulo Cavalcanti Gomes Ferreira^{2*}

¹ Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Diretoria de Pesquisa Científica, Rio de Janeiro, Rio de Janeiro, Brazil, ² Instituto de Bioquímica Médica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil

Abstract

Background: DNA barcoding has been successfully established in animals as a tool for organismal identification and taxonomic clarification. Slower nucleotide substitution rates in plant genomes have made the selection of a DNA barcode for land plants a much more difficult task. The Plant Working Group of the Consortium for the Barcode of Life (CBOL) recommended the two-marker combination *rbcL/matK* as a pragmatic solution to a complex trade-off between universality, sequence quality, discrimination, and cost.

Methodology/Principal Findings: It is expected that a system based on any one, or a small number of plastid genes will fail within certain taxonomic groups with low amounts of plastid variation, while performing well in others. We tested the effectiveness of the proposed CBOL Plant Working Group barcoding markers for land plants in identifying 46 bromeliad species, a group rich in endemic species from the endangered Brazilian Atlantic Rainforest. Although we obtained high quality sequences with the suggested primers, species discrimination in our data set was only 43.48%. Addition of a third marker, *trnH-psbA*, did not show significant improvement. This species identification failure in Bromeliaceae could also be seen in the analysis of the GenBank's *matK* data set. Bromeliaceae's sequence divergence was almost three times lower than the observed for Asteraceae and Orchidaceae. This low variation rate also resulted in poorly resolved tree topologies. Among the three Bromeliaceae subfamilies sampled, Tillandsioideae was the only one recovered as a monophyletic group with high bootstrap value (98.6%). Species paraphyly was a common feature in our sampling.

Conclusions/Significance: Our results show that although DNA barcoding is an important tool for biodiversity assessment, it tends to fail in taxonomy complicated and recently diverged plant groups, such as Bromeliaceae. Additional research might be needed to develop markers capable to discriminate species in these complex botanical groups.

Citation: Maia VH, Mata CSd, Franco LO, Cardoso MA, Cardoso SRS, et al. (2012) DNA Barcoding Bromeliaceae: Achievements and Pitfalls. PLoS ONE 7(1): e29877. doi:10.1371/journal.pone.0029877

Editor: Dirk Steinke, Biodiversity Institute of Ontario - University of Guelph, Canada

Received: July 12, 2011; **Accepted:** December 6, 2011; **Published:** January 12, 2012

Copyright: © 2012 Maia et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: PCGF and ASH received support from Conselho Nacional de Desenvolvimento Científico e Tecnológico research grants. V.H.M is indebted to CNPq for graduate fellowships. The authors are also thankful to Natura for financial support. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Some funding to this work was provided by the commercial company Nature. However, this funding was approved by the Minister of Culture under a law that gives tax exceptions to companies that support important collections such as libraries and so on. In our particular case funding was provided to our group to create a DNA bank of plant species of the Atlantic Rain Forest. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

* E-mail: paulof@bioqmed.ufrj.br

These authors contributed equally to this work.

Introduction

Taxonomy is a key science to understand and monitor biodiversity. However, since taxonomists first started classifying organisms, roughly 1.7 million species have been described, representing only a small fraction of the estimated tens of millions extant species on Earth [1]. This descriptive deficit varies widely, not only among different taxa but also between countries. Unfortunately, the richest in biodiversity are the poorest in human and material resources. Moreover, the tropics, which harbor the majority of Earth's species, are potentially threatened by global climate change [2,3]. Human activities are also causing the extinction of species hundreds of times faster than the natural rate of extinction found in the fossil record [1]. With the predicted climate change and no immediate change in human behavior, the biosphere will be drastically modified within just a few decades, with the number of species facing extinction being estimated at one million by 2050 [4].

This urgent need to inventory and manage diversity created a demand for innovative tools to identify species, such as DNA barcoding, proposed by Hebert *et al.* [5]. A DNA barcode is a short, highly variable, standardized, orthologous DNA sequence used to identify species through comparison to a DNA sequence database. It has been most successfully applied to animals using the mitochondrial cytochrome c oxidase I gene. The *COI*, or *coxI* gene, is variable enough to allow discrimination among closely related species, and yet possesses highly conserved regions that can be easily sequenced with standard protocols [5,6].

Because of slower nucleotide substitution rates in plant mitochondrial DNA [7], plastids have become the primary source of DNA barcodes in plants. After evaluating the performance of seven leading candidate plastid DNA regions (*atpF-atpH*, *psbK-psbI*, *trnH-psbA* spacers and *matK*, *rbcL*, *rpoB*, *rpoC1* genes), the Plant Working Group of the Consortium for the Barcode of Life (CBOL) recommended the two-marker

combination *rbcl/matK* as the standard DNA barcode for plants [8].

It is expected that a system based on any one or small number of plastid genes will fail in certain taxonomic groups with low amounts of plastid variation, while performing well in others [9]. The purpose of the current study is to test the effectiveness of the proposed markers in Bromeliaceae, a group rich in endemic species in the endangered Brazilian Atlantic Rainforest.

The Atlantic Rainforest is a hot spot of plant biodiversity, with a large number of endemic species. Most natural populations of these species are comprised of a low number of individuals and it's believed that for every two endangered Brazilian trees, one is found exclusively in the Atlantic Rainforest ecosystem. Because of its the proximity to urban centers, this complex ecosystem, one of the most threatened in the world, has been the subject of environmental degradation by social and economic pressures. During the last 500 years, the geographical range of the Atlantic Rainforest has been reduced from 12% to 1% of the Brazilian territory.

Materials and Methods

Plant Material

A total of 101 individuals, comprising 46 species were sampled from the Dimitri Sucre greenhouse, which holds the Bromeliaceae scientific collection of the Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, in Rio de Janeiro/RJ, Brazil. Most individuals sampled in this study derived from plants collected in scientific expeditions led by Dr. Gustavo Martinelli. Vouchers of each specimen are deposited in the Herbarium (RB) of the Instituto de Pesquisas Jardim Botânico do Rio de Janeiro. Specimens' vouchers and respective GenBank accessions are available in Table S1.

DNA Extraction, Amplification and Sequencing

Young leaves were collected, cleaned, torn by hand when possible and immediately frozen in liquid nitrogen or dried in silica gel. Frozen tissues were disrupted in liquid nitrogen using a mortar and pestle. Dry material was ground in a mixer-mill disruptor (Mixer Mill 301 Confort, Retsch) using stainless steel beads. Total genomic DNA was isolated using modified CTAB protocol. [10]. All DNA extracted is deposited in the DNA Bank of Brazilian Flora Species, Instituto de Pesquisas Jardim Botânico do Rio de Janeiro.

The two proposed plastid barcoding markers for plants, *matK* and *rbcl* [8], were evaluated using the described sets of primers: 390F/1326R for *matK* [11] and 1F/724R for *rbcl* [12]. The noncoding *trnH-psbA* spacer was also tested using primers *psbA3'f/trnHf* described by Kress *et al.*, 2005 – PNAS 102 (23): 8369–8374.

PCR reactions contained 25 ng of DNA template, 1× reaction buffer (10×: 10 mM Tris-HCl pH8.5, 50 mM KCl, 1.5 mM MgCl₂, 0.01% gelatin), 0.2 mM dNTPs, 10 pmol of each primer and 2.5 units of Taq DNA polymerase, resulting in a final volume of 25 μL. Reactions with *matK* primers also contained 4% of the total reaction volume of DMSO. BSA was added in *rbcl* and *trnH-psbA* amplifications for a final concentration of 0.5 μg/μL.

The following PCR profiles were used:

- 1) *matK*: 94°C 5 min; [26 cycles: 94°C 1 min; 48°C 30 sec; 72°C 1 min]; 72°C 7 min;
- 2) *rbcl*: 94°C 5 min; [30 cycles: 94°C 1 min; 50°C 30 sec; 72°C 1 min]; 72°C 7 min;
- 3) *trnH-psbA*: 94°C 5 min; [30 cycles: 94°C 30 sec; 48°C 30 sec; 72°C 30 sec]; 72°C 5 min

PCR products were purified and sequenced at Macrogen Inc., Seoul, South Korea. Sequencing was conducted under BigDye™ Terminator v3.1 cycling conditions. The reacted products were purified using ethanol precipitation and run using an ABI3730XL automatic sequencer.

Data Acquisition from GenBank

matK sequences of Bromeliaceae, Asteraceae and Orchidaceae were downloaded from GenBank. Sequences with ambiguous bases (more than 10 'Ns' in the whole sequence), or those belonging to unnamed species (i.e. sequences with 'sp.' in the species name) were filtered out. Analyses were conducted with the selected sequences, constituting different data sets for each family (see Table S2 for GenBank accessions). All sequences were equally trimmed in order to make comparisons among data sets possible.

Data Analysis

Sequences were assembled and edited in Geneious Pro 5.0.4 software (Biomatters Ltd.). Prior to assembling, sequences were trimmed based on the quality values of the traces, using the Modified-Mott algorithm implemented on the software. Contig quality was assessed by Confidence Mean value, which is the mean of confidence scores for the contig base calls.

Sequence alignments were conducted on MUSCLE (version 3.8.31) [13], using default parameters, and subsequently checked by visual inspection. The optimality criterion GLOCSA [14] was applied to evaluate the best alignment option, using the software GLOCSER (downloaded at goten.ibiologia.unam.mx/glocsa/wp-content/uploads/2010/12/glocser-sept2010.zip).

Phylogenetic analyses were made with the combined data sets, using three different methods: maximum parsimony (MP), distance, and maximum likelihood (ML). *Typha angustifolia* sequences downloaded from GenBank (accessions GQ436382, AY952419 and EU750604) were used as an outgroup. Parsimony analyses were performed using PAUP* (version 4.0b10) [15]. Tree searches were conducted using heuristic search option RANDOM addition sequences (1000 replicates) holding 10 trees per replicate, and tree bisection-reconnection (TBR) branch swapping, with retention of multiple parsimonious trees (MAXTREES was set to 5000). Branch support was given by 1000 bootstrap pseudoreplicates, with the following parameters for tree search: simple taxon addition, TBR branch swapping, and holding 10 trees at each step. Incongruence between marker partitions was evaluated by the Incongruence Length Difference (ILD) test [16].

Distance analysis was carried out on MEGA 5 [17], using the NJ algorithm and Jukes-Cantor as the model of substitution. The choice of this model was based on Nei and Kumar [18] (overall mean: p distance = 0.010 transitions/transversions = 2.09). For the ML analysis, HKY+I+G model was selected as the best-fit model by AIC implemented in jModeltest (version 0.0.1) [19]. PHYML (version 3.0) [20] was used for phylogeny estimate with a BIONJ starting tree [21]. Bootstrap support was accessed by 1000 and 100 pseudoreplicates in NJ and ML analyses, respectively.

MEGA5 was also used to calculate K2P pairwise and overall distances to each one of the previously designated DNA barcode regions. In order to compare *matK* variation in different plant families, we also calculated K2P overall distance to the GenBank's data sets.

A local BLAST search method was performed to test species identification capability, as described previously [22]. This method was applied to all data sets (GenBank, *rbcl*, *matK*, *trnH-psbA* and combinations of regions). In short, a reference library was constructed using the "makeblastDB" command in BLAST+ [23] for each region and combination of regions (*rbcl+trnH-psbA*,

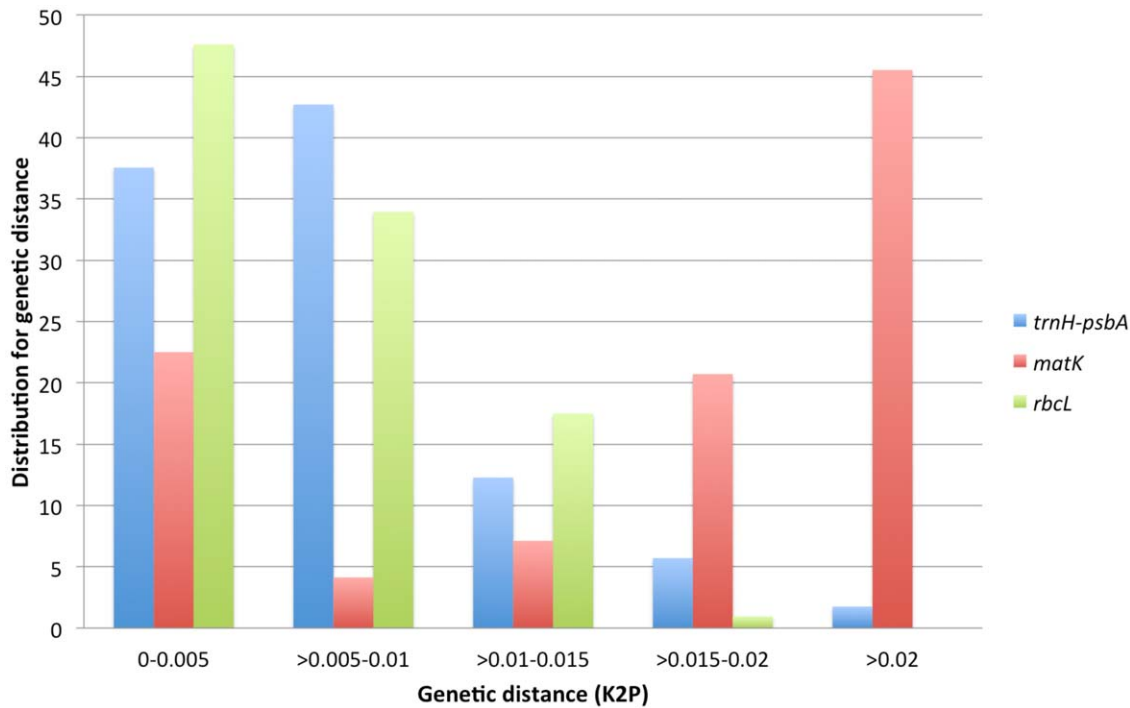


Figure 1. Relative distribution of K2P distances across all sequence pairs of Bromeliaceae data set for different markers.
doi:10.1371/journal.pone.0029877.g001

rbcL+matK, *matK+trnH-psbA* and *rbcL+matK+trnH-psbA*). The barcode sequence of each species was then queried against the reference library with the “blastn” command. Only species matching 100% in sequence similarity over the entire sequence length with their own reference sequence were counted as successful identifications; those that also matched 100% over the entire sequence length but had the reference sequence of one or more different species were counted as failures.

All statistical analyses regarding sequence quality, PCR/sequencing success, and species discrimination among markers were conducted on STATISTICA software (version 7; StatSoft Inc., 2004), using the test of Tukey at $P = 0.05$.

Results

Success rates for PCR/sequencing and Sequence Quality

matK sequences were obtained from all 101 DNA samples, while *rbcL* and *trnH-psbA* sequences could only be obtained from 92 and 99 samples, respectively. Polymerase chain reactions (PCR) of the three regions were always successful, but sequencing failed for these missing samples. PCR/sequencing success was statistically different only for *rbcL* (Tukey test; $P = 0.05$). The average sequence quality values (Confidence Mean scores) were 46.6, 46.5 and 43.1

for *rbcL*, *trnH-psbA* and *matK*, respectively. Although *matK* presented a satisfactory sequence quality, it was significantly different from the other two markers.

Distance and Phylogenetic Analyses

Sequences from *rbcL* and *matK* data sets resulted in 632 bp and 797 bp length alignments, respectively, with no indels. The *trnH-psbA* aligned matrix was 679 bp long with 29 indels, ranging from 1 to 38 bp (mean size = 11.4). Possible adjustments on *trnH-psbA* alignment were not done since MUSCLE generated the alignment with the best GLOCSA score (815.16). Sequences without complete data set were excluded from combined analysis.

Figure 1 shows the distribution of interspecific K2P distances among all sequences that were successfully obtained for *matK*, *rbcL* and *trnH-psbA*.

Estimates of average evolutionary divergence over all GenBank’s *matK* data set for Bromeliaceae, Asteraceae and Orchidaceae, using K2P model, are shown in the Table 1.

Phylogenetic Analyses

For the phylogenetic analyses, the combined data set was used since ILD test showed no significant difference among markers

Table 1. Comparison of identification efficiency of *matK* region among GenBank’s data sets, representing three angiosperm families.

Families	N° Sequences	N° Species	Success Rates (Blast Test)	Overall K2P Distance
Asteraceae	778	577	74.87%	4.8%
Bromeliaceae	422	334	52.40%	1.9%
Orchidaceae	365	297	64.67%	6.2%

doi:10.1371/journal.pone.0029877.t001

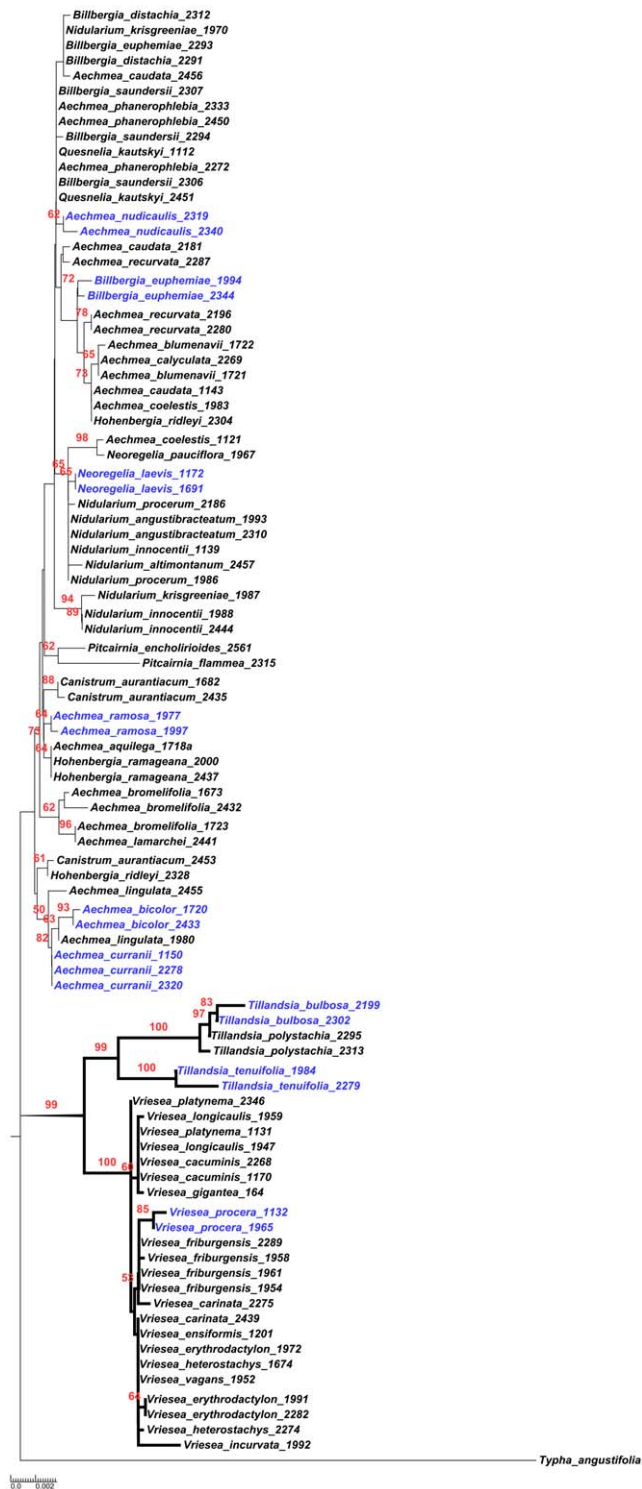


Figure 2. NJ tree of genetic distance (JC) for 91 Bromeliaceae specimens, based on *rbcL*, *matK* and *trnH-psbA* combined analysis. Genera and species recovered as monophyletic are highlighted (thicker branches and names in blue, respectively). Numbers above branches correspond to bootstrap support $\geq 50\%$. doi:10.1371/journal.pone.0029877.g002

($P = 0.082$). All three phylogenetic reconstruction methods recovered very similar and poorly resolved topologies. MP analysis resulted in 5000 equally parsimonious trees ($L = 318$). The NJ tree

is shown in Figure 2. Of the 35 species with multiple accessions, only six were recovered as monophyletic in all analyses.

Blast Analysis

Summaries of species discrimination rates based on the three regions and their combination are shown in Table 2. The single region with highest success rate is *matK* (27.66%). The combination of the three regions increases success rate to 44.44%. Statistical significance among markers and their combinations is shown in Table S3.

Blast test performed with GenBank's data sets (Table 1) exhibited the lowest discrimination rate for Bromeliaceae (52.40%).

Discussion

DNA barcoding has helped to rejuvenate taxonomy, a key science to understanding and monitoring biodiversity. Estimates of regional species diversity, abundances, and geographic distributions are essential to comprehend how species-rich communities are assembled and how they might be conserved.

In this study, we tested the effectiveness of the proposed CBOL Plant Working Group barcoding *markers* for land plants in identifying 46 bromeliad species. Bromeliaceae form a large, morphologically distinctive, ecologically diverse and species-rich family of angiosperms, native to the tropics and subtropics of the New World [24]. Many species are endemic to the Atlantic Rainforest, one of the most threatened tropical ecosystems in the world.

The *rbcL/matK* two-marker combination was proposed as the standard plant barcode for its universality, sequence quality, discrimination, and costs. Although we obtained high quality sequences with the suggested primers, species discrimination in our data set was 43.48%; much lower than the 72% success found by CBOL Plant Working Group. The authors pointed out that further resolution could be achieved with the use of supplementary loci, as non-coding plastid regions, and internal transcribed spacers of nuclear ribosomal DNA, when direct sequencing of this locus is possible. Among the non-coding plastid regions, *trnH-psbA* spacer remains the leading candidate as a source of additional data [8]. In our study, as seen in Table 2, the addition of a third marker did not show significant improvement, raising the identification rate only to 44.44%.

matK and *trnH-psbA* showed the highest individual identification rates (Table 2), confirming marker's previously reported species discrimination power. Improvements in *matK* primers' universality and in obtaining high quality *trnH-psbA* bidirectional sequences were not necessary. The suggested *matK* primers performed perfectly for Bromeliaceae. All samples were successfully amplified and sequenced using a single primer pair. Likewise, only two out of 101 samples failed in generating high quality *trnH-psbA* bidirectional sequences after the second trial. Although *rbcL* showed the lowest species identification rate in Bromeliaceae, its sequence variation (overall K2P distance) was similar to the obtained for *trnH-psbA* (Table 2). *matK* sequence variation was twice the amount obtained for the two other markers (Table 2). This trend can also be observed in the pairwise K2P distance distribution frequencies (Figure 1). *matK/trnH-psbA* are accounted for the second highest two-locus combination resolution (34.78%), while *rbcL/trnH-psbA* almost maintained *trnH-psbA* success identification rate, 26.67% and 26.09%, respectively (Table 2).

In Rosaceae, the combination *rbcL/matK* correctly identified 96% of the sampled species [25]. Other families such as Fabaceae [26], Euphorbiaceae [27], and Asteraceae [28] present success

Table 2. Comparison of identification efficiency of the plastid regions and their combinations in Bromeliaceae data set.

Region	N° Species (Correct Identification)	Success Rates (Blast Test)	Overall K2P Distance
<i>rbcl</i>	47 (9)	19.15%	0.6%
<i>matK</i>	47 (13)	27.66%	1.4%
<i>trnH-psbA</i>	46 (12)	26.09%	0.7%
<i>rbcl+matK</i>	46 (20)	43.48%	-
<i>rbcl+trnH-psbA</i>	45 (12)	26.67%	-
<i>matK+trnH-psbA</i>	46 (16)	34.78%	-
<i>rbcl+matK+trnH-psbA</i>	45 (20)	44.44%	-

doi:10.1371/journal.pone.0029877.t002

identification rates over 80%, using *matK* as a single marker. In our Bromeliaceae data set, even the three regions combined yielded a low identification rate (Table 2).

This species identification failure in Bromeliaceae could also be seen in the analysis of GenBank's *matK* data set. Bromeliaceae exhibited the lowest success rate compared to Asteraceae and Orchidaceae (Table 1). These values seem to be related to their sequence variation (overall K2P distance), since Bromeliaceae sequence divergence was almost three times lower than the observed for the other families.

This low variation rate seen in Bromeliaceae also resulted in poorly resolved tree topologies, regardless of the method that was used to infer the phylogeny. Of the three subfamilies represented in our study, only Tillandsioideae was recovered as monophyletic with a high bootstrap value (98.6%). The monophyly of its two genera, *Tillandsia* and *Vriesia*, was highly supported. None of the other genera from other subfamilies were recovered as monophyletic. In fact, paraphyly has been reported for several Bromeliaceae genera [24]. Species paraphyly was also a common feature in our sampling; only ten were recovered monophyletic (Figure 2). These results suggest that, in the Bromeliaceae, taxon definition below genera may be quite artificial; consequence of a rapid morphological divergence contrasting with a high genetic conservatism.

The lack of barcoding gaps, evidenced by species paraphyly and low identification rates, may be explained by unclear species boundaries. Taxonomy of the group is complicated by interspecific hybridization, a mixture of sexual and asexual reproduction, and recent species divergence, as also seen in wild potatoes [29]. Recent lineages with a great number of species tend to have lots of close relatives, reducing levels of interspecific divergence [30]. Bromeliaceae is an example of recent radiation. Extant subfamilies began to diverge from each other only about 19 Mya [24].

DNA barcoding is an important tool for biodiversity assessment. However, it tends to fail in taxonomy complicated and recently diverged groups, such as Bromeliaceae. The use of sequences from multiple genetic sources may improve this scenario. Next-

generation sequencing technologies may provide genomic information for comparisons within and among multiple taxonomic levels [31]. Nevertheless, although limited, generated barcoding data reflects the evolution of the group and should not be ignored.

Supporting Information

Table S1 Accessions used in the work. All specimens' vouchers are deposited in the herbarium of Rio de Janeiro Botanical Garden (RB). (DOC)

Table S2 Accession codes of *matK* sequences downloaded from GenBank, belonging to three angiosperm families. (XLS)

Table S3 Statistical significance of species discrimination between markers and markers combinations. Significant comparisons are in red ($P=0.05$, Tukey test); $r = rbcl$, $m = matK$ and $t = trnH-psbA$. (DOC)

Acknowledgments

We thank Dr. Gustavo Martinelli for providing plant material and information about the species; Rodolfo C. R. de Abreu for his kind assistance in statistical analyses; Heron Zanellatto, Nara Lucia C. Vasconcellos and Sr. Mourão for organizing and maintaining the Dimitri Sucre greenhouse scientific Bromeliaceae collection and for confirming material identification.

Author Contributions

Conceived and designed the experiments: LOF MAC SRCS ASH PCGF. Performed the experiments: CSdM VHM LOF. Analyzed the data: VHM CSdM LOF PCGF. Contributed reagents/materials/analysis tools: CSdM VHM LOF PCF. Wrote the paper: CSdM VHM LOF ASH SRCS MAC PCGF.

References

- Vernooij R, Haribabu E, Muller MR, Vogel JH, Hebert PDN, et al. (2010) Barcoding life to conserve biological diversity: beyond the taxonomic imperative. *Plos Biol* 8: e1000417. doi:10.1371/journal.pbio.1000417.
- Bradshaw CJA, Sodhi NS, Brook BW (2009) Tropical turmoil: a biodiversity tragedy in progress. *Front Ecol Environ* 7: 79–87. doi:10.1890/070193.
- Malhi Y, Roberts JT, Betts RA, Killeen TJ, Li W, et al. (2008) Climate change, deforestation, and the fate of the Amazon. *Science* 319: 169–172. doi:10.1126/science.1146961.
- Savolainen V, Reeves G (2004) A plea for DNA banking. *Science* 304: 1445–1445. doi:10.1126/science.304.5676.1445b.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proc Biol Sci* 270: 313–321. doi:10.1098/rspb.2002.2218.
- Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PDN (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *P Natl Acad Sci Usa* 103: 3657–3662. doi:10.1073/pnas.0511318103.
- Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD (2007) Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evol Biol* 7: 135. doi:10.1186/1471-2148-7-135.
- CBOL Plant Working Group (2009) A DNA barcode for land plants. *P Natl Acad Sci Usa* 106: 12794–12797. doi:10.1073/pnas.0905845106.
- Steven GN, Subramanyam R (2009) Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). *Molecular Ecology Resources* 9 Suppl s1: 172–180. doi:10.1111/j.1755-0998.2009.02642.x.

10. Dolye JJ, Doyle JL (1987) A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. *Phytochemical Bulletin* 9: 11–15.
11. Cuénoud P, Savolainen V, Chatrou L, Powell M, Grayer R, et al. (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *Am J Bot* 89: 132–144.
12. Fay M, Cameron K, Prance G, Lledó M (1997) Familial relationships of *Rhabdodendron* (Rhabdodendraceae): plastid *rbcL* sequences indicate a caryophyllid placement. *Kew Bulletin* 52(4): 923–932.
13. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5): 1792–97.
14. Arenas-Diaz E, Ochoterena H, Rodriguez-Vazquez K (2009) Multiple sequence alignment using a genetic algorithm and GLOCSA. *Journal of Artificial Evolution and Applications* 2009: 963150.
15. Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). ;Version 4. Sinauer Associates, Sunderland, Massachusetts.
16. Farris JS, Källersjö M, Kluge AG, Bult C (1994) Testing Significance of Incongruence. *Cladistics* 10(3): 315–319.
17. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol*; doi:10.1093/molbev/msr121.
18. Nei M, Kumar S (2000) *Molecular evolution and phylogenetics*. Oxford University Press, New York, NY, USA.
19. Posada D (2008) jModelTest: Phylogenetic Model Averaging. *Mol Biol Evol* 25(7): 1253–1256.
20. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52(5): 696–704.
21. Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol* 14(7): 685–95.
22. Ebihara A, Nitta JH, Ito M (2010) Molecular Species Identification with Rich Floristic Sampling: DNA Barcoding the Pteridophyte Flora of Japan. *Plos One* 5: e15136. doi:10.1371/journal.pone.0015136.
23. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. (2009) BLAST+: architecture and applications. *BMC Bioinf* 10: 421.
24. Givnish TJ, Barfuss MHJ, Ee BV, Riina R, Schulte K, et al. (2011) Phylogeny, adaptive radiation, and historical biogeography in Bromeliaceae: Insights from an eight-locus plastid phylogeny. *Am J Bot* 98: 872–895. doi:10.3732/ajb.1000059.
25. Pang X, Song J, Zhu Y, Xu H, Huang L, et al. (2011) Applying plant DNA barcodes for Rosaceae species identification. *Cladistics* 27: 165–170. doi:10.1111/j.1096-0031.2010.00328.x.
26. Gao T, Sun Z, Yao H, Song J, Zhu Y, et al. (2011) Identification of Fabaceae plants using the DNA barcode *matK*. *Planta Med* 77: 92–94. doi:10.1055/s-0030-1250050.
27. Pang X, Song J, Zhu Y, Xie C, Chen S (2010) Using DNA barcoding to identify species within Euphorbiaceae. *Planta Med* 76: 1784–1786. doi:10.1055/s-0030-1249806.
28. Gao T, Yao H, Song J, Zhu Y, Liu C, et al. (2010) Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. *BMC Evol Biol* 10: 324. doi:10.1186/1471-2148-10-324.
29. Spooner DM (2009) DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. *Am J Bot* 96: 1177–1189. doi:10.3732/ajb.0800246.
30. Gonzalez MA, Baraloto C, Engel J, Mori SA, Petronelli P, et al. (2009) Identification of Amazonian Trees with DNA Barcodes. *Plos One* 4: e7483. doi:10.1371/journal.pone.0007483.
31. Steele PR, Pires JC (2011) Biodiversity assessment: State-of-the-art techniques in phylogenomics and species identification. *Am J Bot* 98: 415–425. doi:10.3732/ajb.1000296.