rbcL and *matK* Earn Two Thumbs Up as the Core DNA Barcode for Ferns

Fay-Wei Li[1]*, Li-Yaung Kuo[2], Carl J. Rothfels[1], Atsushi Ebihara[3], Wen-Liang Chiou[4], Michael D. Windham[1], Kathleen M. Pryer[1]

1 Department of Biology, Duke University, Durham, North Carolina, United States of America, **2** Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei, Taiwan, **3** Department of Botany, National Museum of Nature and Science, Tsukuba-shi, Ibaraki, Japan, **4** Division of Botanical Garden, Taiwan Forestry Research Institute, Taipei, Taiwan

Abstract

Background: DNA barcoding will revolutionize our understanding of fern ecology, most especially because the accurate identification of the independent but cryptic gametophyte phase of the fern's life history—an endeavor previously impossible—will finally be feasible. In this study, we assess the discriminatory power of the core plant DNA barcode (*rbcL* and *matK*), as well as alternatively proposed fern barcodes (*trnH-psbA* and *trnL-F*), across all major fern lineages. We also present plastid barcode data for two genera in the hyperdiverse polypod clade—*Deparia* (Woodsiaceae) and the *Cheilanthes marginata* group (currently being segregated as a new genus of Pteridaceae)—to further evaluate the resolving power of these loci.

Principal Findings: Our results clearly demonstrate the value of *matK* data, previously unavailable in ferns because of difficulties in amplification due to a major rearrangement of the plastid genome. With its high sequence variation, *matK* complements *rbcL* to provide a two-locus barcode with strong resolving power. With sequence variation comparable to *matK*, *trnL-F* appears to be a suitable alternative barcode region in ferns, and perhaps should be added to the core barcode region if universal primer development for *matK* fails. In contrast, *trnH-psbA* shows dramatically reduced sequence variation for the majority of ferns. This is likely due to the translocation of this segment of the plastid genome into the inverted repeat regions, which are known to have a highly constrained substitution rate.

Conclusions: Our study provides the first endorsement of the two-locus barcode (*rbcL+matK*) in ferns, and favors *trnL-F* over *trnH-psbA* as a potential back-up locus. Future work should focus on gathering more fern *matK* sequence data to facilitate universal primer development.

Citation: Li F-W, Kuo L-Y, Rothfels CJ, Ebihara A, Chiou W-L, et al. (2011) *rbcL* and *matK* Earn Two Thumbs Up as the Core DNA Barcode for Ferns. PLoS ONE 6(10): e26597. doi:10.1371/journal.pone.0026597

Editor: Sergios-Orestis Kolokotronis, Barnard College, Columbia University, United States of America

Received June 3, 2011; **Accepted** September 29, 2011; **Published** October 20, 2011

Funding: A National Science Foundation (www.nsf.gov) grant #DEB-0717398 to KMP and MDW (co-authors). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: fl43@duke.edu

Introduction

In all land plants—from bryophytes to angiosperms—the typical sexual life cycle involves the alternation of a diploid sporophyte phase with a haploid gametophyte phase. Ferns and lycophytes are unique among land plants in that both sporophyte and gametophyte are not only visible to the unaided eye, but they are completely independent from one another [1]. Although diminutive and inconspicuous, fern gametophytes are key players in fern ecology and biogeography: many are thought to have wider geographic distributions than their sporophytic counterparts [2–4], some can exist indefinitely without producing sporophytes [5,6], and others may be involved in hybridization events far outside the range of the parental sporophyte [7]. However, because the morphology of fern gametophytes is so reduced, it has been very difficult to confidently assign gametophytes to species, or, frequently, even to particular genera. As a result, ecological studies of gametophytes have largely been restricted to temperate regions where relatively few species exist [8], or to well-studied biological field stations in the tropics [9].

DNA barcoding offers a possible solution to this problem and could greatly improve our understanding of fern gametophytes and their biology. Recently, unknown fern gametophytes were shown to be identifiable, often to species level, by using plastid DNA sequences [5,10,11], suggesting that this DNA-based identification tool has the potential to be applied to large-scale ecological surveys [12]. The DNA barcoding approach has also been useful in distinguishing among fern species in the horticultural trade [13] and in Chinese herbal medicine [14,15], two areas where species names are frequently confused. Despite these promising applications, ferns, with their critical phylogenetic position as sister to seed plants, have largely been neglected in choosing the standardized barcode for all land plants [16,17].

Recently, the Consortium for the Barcode of Life (CBOL) approved two plastid loci, *rbcL* and *matK*, as the official DNA barcode for all land plants [18,19], while urging further data collection on *trnH-psbA* to assess its potential to be added to the land plant barcode. CBOL's pronouncement posed a serious challenge for fern systematists and ecologists because *matK* had been recovered from only one fern species in the previous loci evaluation studies [20–22]. Because of the difficulties involved in obtaining *matK* data for ferns, Ebihara et al. [23] and de Groot et al. [11] proposed *trnH-psbA* and *trnL-F*, respectively, as possible substitutes.

In most plants, *matK* is nested within a *trnK* intron in the large single copy region of the plastid genome and can be amplified using primers targeting the flanking *trnK* exons [24,25]; as these full-length *matK* sequences accumulate, further primer development for *matK* coding region should be relatively easy. In most ferns, however, the flanking *trnK* exons are absent [26–28], and the amplification of full-length *matK* is very difficult, thereby hindering primer development. Only recently has novel primer design helped to overcome this obstacle, with *matK* sequences now available for representatives from all fern families (*sensu* [29]) [28]. The primary aim of our study is to provide a broad overview of sequence variation across fern lineages for the core DNA barcode (*rbcL* and *matK*), as well as for the two alternatively proposed barcode regions (*trnH-psbA* and *trnL-F*). We then focus particular attention on two genera within the hyperdiverse polypod clade—*Deparia* (Woodsiaceae) and the *Cheilanthes marginata* group (a group of 17 species currently being segregated as a new genus of Pteridaceae; F.W. Li et al., unpublished). These case studies provide more detailed information regarding the resolving power of all four loci for species level identifications.

Results

Of the four plastid loci examined in the large-scale comparisons, *trnL-F* is the most variable across ferns, followed by *matK*, *trnH-psbA*, and then *rbcL* (Fig. 1; $p<0.0001$ for each comparison in Wilcoxon matched-pairs signed-rank tests, after Bonferroni correction for multiple comparisons). In contrast to its high levels of variation in most other plant lineages, *trnH-psbA* shows a markedly reduced variability in ferns (Table 1), such that 99.1% of the species pairs tested show lower divergence at *trnH-psbA* compared to *matK* (Fig. 1B), and only 5.6% are more than twice as variable as *rbcL* (Fig. 1C). This reduced variation in *trnH-psbA* is most pronounced in the recent-diverging fern lineages Cyatheales and Polypodiales (Fig. 1B–C, 1F), which together account for almost 90% of fern diversity [1].

Pairwise sequence divergence within the focal polypod genus *Deparia* is mostly comparable to the large-scale trends observed across all ferns (cf. Fig. 1A–B with Fig. 2A–B, and Fig. 1D–F with Fig. 2D–F), although *trnH-psbA* is even more conservative (cf. Fig. 1C, 1F with Fig. 2C, 2F) (Table 1). Although we do not have *trnL-F* data for the *Cheilanthes marginata* group, it shows the same general trends observed in *Deparia* for *matK*, *rbcL* and *trnH-psbA* (Fig. 2A–C). The average sequence divergence for *trnH-psbA* is lower than for all other loci tested (Table 1), and 83.2% and 58.5% of the species pairs in the *C. marginata* group and *Deparia*, respectively, exhibit a *trnH-psbA* divergence that is lower than for *rbcL* (Fig. 2C).

Finally, we examined the discriminating power for each locus and locus combination within the *C. marginata* group and within *Deparia*. Table 2 shows that among the single-locus barcodes, *matK* has the highest success rate in species discrimination and *trnH-psbA* the lowest. More importantly, when considering each locus and all combinations of loci, the highest success rate is provided by *matK*+*rbcL*, the official core DNA barcode [18,19,22], while including additional *trnH-psbA* did not increase the rate (Table 2).

Discussion

Toward A Consensus Barcode For Ferns

Two proposals regarding a global DNA barcode for all land plants were recently formulated and presented to CBOL [18,19]. One consisted of *rbcL* and *matK* while the other included *rbcL*, *matK* and *trnH-psbA*. CBOL officially approved the *rbcL*+*matK* combination, and encouraged more data collection on *trnH-psbA* to assess its potential as a backup barcoding locus [18]. Because *matK* had been previously thought to be unattainable in ferns, two non-coding loci, *trnH-psbA* and *trnL-F*, were independently proposed as alternative barcoding loci [11,23].

In this study, we provide the first thorough evaluation of the official CBOL land plant barcode (*matK* and *rbcL*) and the two alternative (*trnH-psbA* and *trnL-F*) loci for ferns. Our results build on the recent demonstration that *matK* is attainable in ferns [28], and shows that its variability is consistently high across fern lineages. Even within the species complexes represented by our two focal polypod genera, *matK* and *rbcL* together provide the highest discriminating power, supporting their use as the official core DNA barcode. It should be noted that our *matK* and *rbcL* sequences are longer than the barcode regions specified by CBOL [19]; however, we do not believe this affects our conclusions. Although universal *matK* primers remain elusive in ferns, we believe primer development will be considerably improved as more sequences become available. Therefore, it would be biased at this stage to examine and compare PCR amplification rate and sequence quality against other loci. However, if attempts to design universal *matK* primers eventually fail, our results suggest that *trnL-F* would be a good alternative locus since variation within *trnL-F* across ferns is comparable to that observed in *matK*.

On the other hand, we find *trnH-psbA* to be an unsuitable barcode for the majority of ferns, despite its obvious utility in seed plants [20,30–35], mosses [36], and the early-diverging fern order, Hymenophyllales [37] (Table 1). Our results indicate that the nucleotide substitution rate for *trnH-psbA* is greatly reduced, especially in two recent-diverging lineages (Cyatheales and Polypodiales) that together comprise nearly 90% of fern diversity. This reduced variation in the recent-diverging ferns was also reported by Ebihara et al. [23], but it was not considered likely to be a major drawback for barcoding. However, data from our two focal polypod genera reveal that the ability of *trnH-psbA* to discriminate species is the lowest among the loci we tested. Considering the limited usefulness of *trnH-psbA* in ferns, we recommend adoption of the official CBOL land plant barcode (*rbcL*+*matK*) for future fern studies.

An Unexpected Substitution Rate Reduction in *trnH-psbA* in Ferns

Our data provide evidence of an abrupt reduction in *trnH-psbA* sequence variation within ferns (Figure 1B–C, 1F, 2B–C, 2F; Table 1). This apparent deceleration in substitution rate seems to occur on the same branch of the fern phylogeny where the translocation of *trnH-psbA* into the inverted repeat (IR) region of the plastid genome is predicted to have occurred [38]: on the branch leading to Schizaeales, Salviniales, Cyatheales and Polypodiales (Fig. 1G, arrow). The IR region comprises two identical copies of the plastid genome that are separated by the large- and small-single copy regions (the LSR and SSR, respectively), and nucleotide substitution rates in the IR region have been shown to be dramatically slower than in either the LSR or SSR [39–42].

Lower substitution rates in the IR region (relative to the rest of the plastid genome) were originally attributed to the predomi-

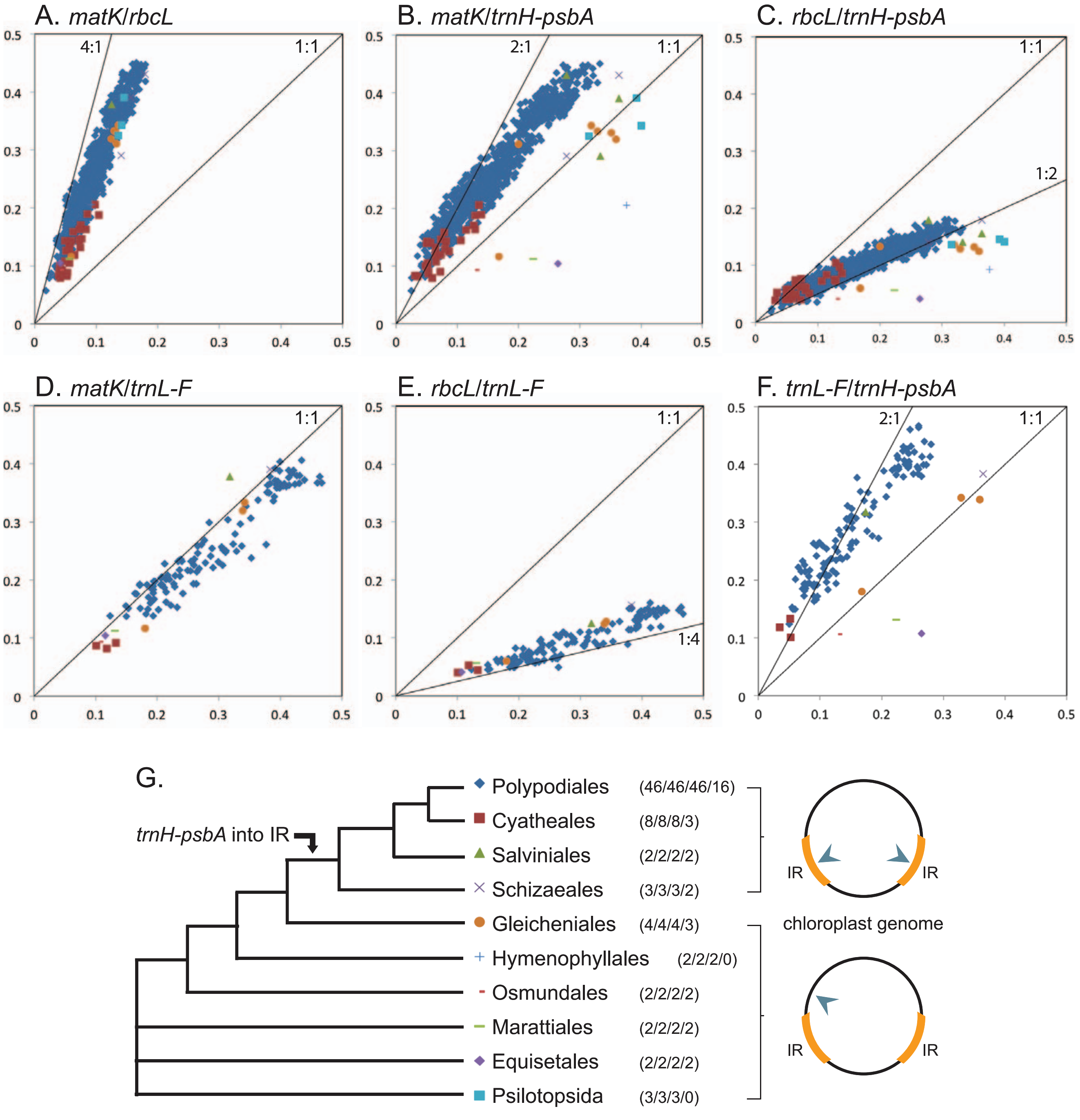

Figure 1. Large-scale loci comparisons across ferns. The x and y axes depict the *p*-distances calculated for each species pair within each fern order (Psilotales and Ophioglossales are combined here into class Psilotopsida). All loci comparisons are presented as y-axis vs x-axis: (A) *matK* vs *rbcL*, (B) *matK* vs *trnH-psbA*, (C) *rbcL* vs *trnH-psbA*, (D) *matK* vs *trnL-F*, (E) *rbcL* vs *trnL-F*, (F) *trnL-F* vs *trnH-psbA*. The lines in each panel (labeled with the ratios 1:1, 1:2, 1:4, 4:1 or 2:1) are not regression lines, but are drawn to guide the eye in interpreting the results. (G) Phylogenetic relationships among fern orders, taxonomic symbols, and number of species compared per order for each locus (*matK*, *rbcL*, *trnH-psbA*, *trnL-F*, respectively). The arrow on the phylogenetic topology points to the branch where the *trnH-psbA* region of the plastid genome is predicted to have been translocated to the inverted repeat region [38]. The arrowheads point to the locations of *trnH-psbA* in the plastid genome.
doi:10.1371/journal.pone.0026597.g001

nance of conservative rRNA genes in this region [41]; however, it has since been shown that rates are ubiquitously slow in the IR region regardless of rRNA content [42]. Two additional hypotheses have been put forth to address this rate disparity—a reduced mutation rate in the IR region, or biased gene conversion between the repeats that tend to correct mutations back to the wild-type states [42]. Invoking a reduced mutation rate may be unnecessary since it can be caused by biased gene conversion and it has been theoretically determined that a slight conversion bias could explain a reduced substitution rate [43]. An empirical study on legumes demonstrated support for this idea, showing that genes that were typically located in the IR region

Table 1. Sequence variation comparisons within different plant groups[1].

	Cheilanthes marginata group (recent-diverging ferns)^2	*Deparia* (recent-diverging ferns)^2	Polypodiales (recent-diverging ferns)^2	Cyatheales (recent-diverging ferns)^2	Hymenophyllales (early-diverging ferns)^2 [37]	Mosses^3 [36]	*Quercus* (Fagaceae)^3 [35]	*Alnus* (Betulaceae)^3 [32]	*Berberis* (Berberidaceae)^3 [31]	*Acacia* (Fabaceae)^2 [33]	Myristicaceae^2 [34]	Angiosperms^3 [16]	Land plants^2 [20]
matK	0.0243	0.0362	0.2667	0.1300	-	-	0.0030	0.0093	0.0050	0.0150	0.0420	0.0125	0.0113
rbcL	0.0099	0.0129	0.0974	0.0614	0.0358	0.0621	0.0010	0.0018	0.0010	0.0140	0.0020	0.0079	0.0129^4
trnH-psbA	0.0076	0.0117	0.1632	0.0769	0.1900	0.1243	0.0103	0.0210	0.0090	0.2010	0.0600	0.0216	0.0269
trnL-F	-	0.0348	-	-	-	0.0973	-	-	-	-	-	-	-

This table does not intend to comprehensively cover the literature, but rather to represent a broad phylogenetic range.

[1] loci variations should only be compared within each plant group; comparison among groups is not valid.

[2] uncorrected *p*-distance.

[3] K2P distance.

[4] *rbcL*-a: a 550–600 bp subset of *rbcL*.

doi:10.1371/journal.pone.0026597.t001

showed an accelerated substitution rate when the IR structure disappeared [44]. Our results from fern *trnH-psbA* provide further evidence, but from the opposite perspective—there is an apparent deceleration in substitution rate when genes are translocated into the IR region. Future statistical analyses, as well as an investigation of *ycf2*, another gene that was co-translocated with *trnH-psbA* into the IR region [38], should better characterize the dynamics of plastid genome evolution in ferns. It is nevertheless evident that because of its low substitution rate in the majority of ferns, *trnH-psbA* is not a suitable DNA barcode region for the fern lineage as a whole.

Low Discriminating Power Within Species Complexes

Because the current DNA barcoding approach in plants relies solely on plastid loci that are mostly uniparentally inherited, it is expected that barcoding will not work well within species complexes where hybridization and polyploidy are frequent [19,31,45,46]. Of our two genus case studies, the *Cheilanthes marginata* group provides a clear illustration of the problem. In addition to a series of diploid species that are easily distinguished by the official *rbcL+matK* barcode, the *C. marginata* group includes two species complexes, each composed of four morphological species that are polyploids of unknown origin (F.W. Li et al., unpublished) [47]. DNA barcoding only discriminates one species in the *C. angustifolia* complex, and none in the *C. marginata* complex (Supporting Information Table S1), thus recognizing less than half of the species-level biodiversity predicted on the basis of morphology. A comparable lack of discriminating power was also reported within species complexes in barcoding studies of Japanese [23] and north-western European ferns [11].

One might argue that incorporating a nuclear locus (e.g., the internal transcribed spacer (ITS) region) could better solve the species complex problem. However, we are hesitant to recommend the use of nuclear loci in ferns (where polyploidy is frequent), because cloning is usually required, not only to obtain clear sequencing results but also to acquire all possible copies. In addition, although ITS has been shown to have high discriminating power in certain plants and animals [48,49], these results needs to be interpreted with caution in ferns, where most of the ITS sequences reported for ferns to date [48–50] are nearly identical to ITS sequences reported for angiosperms (e.g., Asteraceae, Apiaceae or Fabaceae based on BLAST searches in April 2011). It should be noted that disentangling species complexes, which requires extensive genetic and chromosomal analyses, is beyond the expected goals of DNA barcoding. As shown here, the official CBOL land plant barcode allows the identification of most species derived through divergent evolution (as opposed to recent reticulate evolution), and such resolution should be sufficient for most applications [19].

Conclusion

By incorporating both large-scale analyses and genus-level case studies, our study represents the first thorough evaluation of the official CBOL land plant barcode (*matK* and *rbcL*), as well as of two alternative barcode loci (*trnH-psbA* and *trnL-F*), for ferns. Our results provide a strong endorsement of the two-locus barcode (*rbcL+matK*) in ferns, and favor *trnL-F* over *trnH-psbA* as a potential back-up locus. The dramatically reduced variation observed in *trnH-psbA* is likely due to its translocation into the IR region of the plastid genome. Future work should focus on gathering more *matK* sequences for improved primer development, as well as examining PCR amplification and sequencing quality.

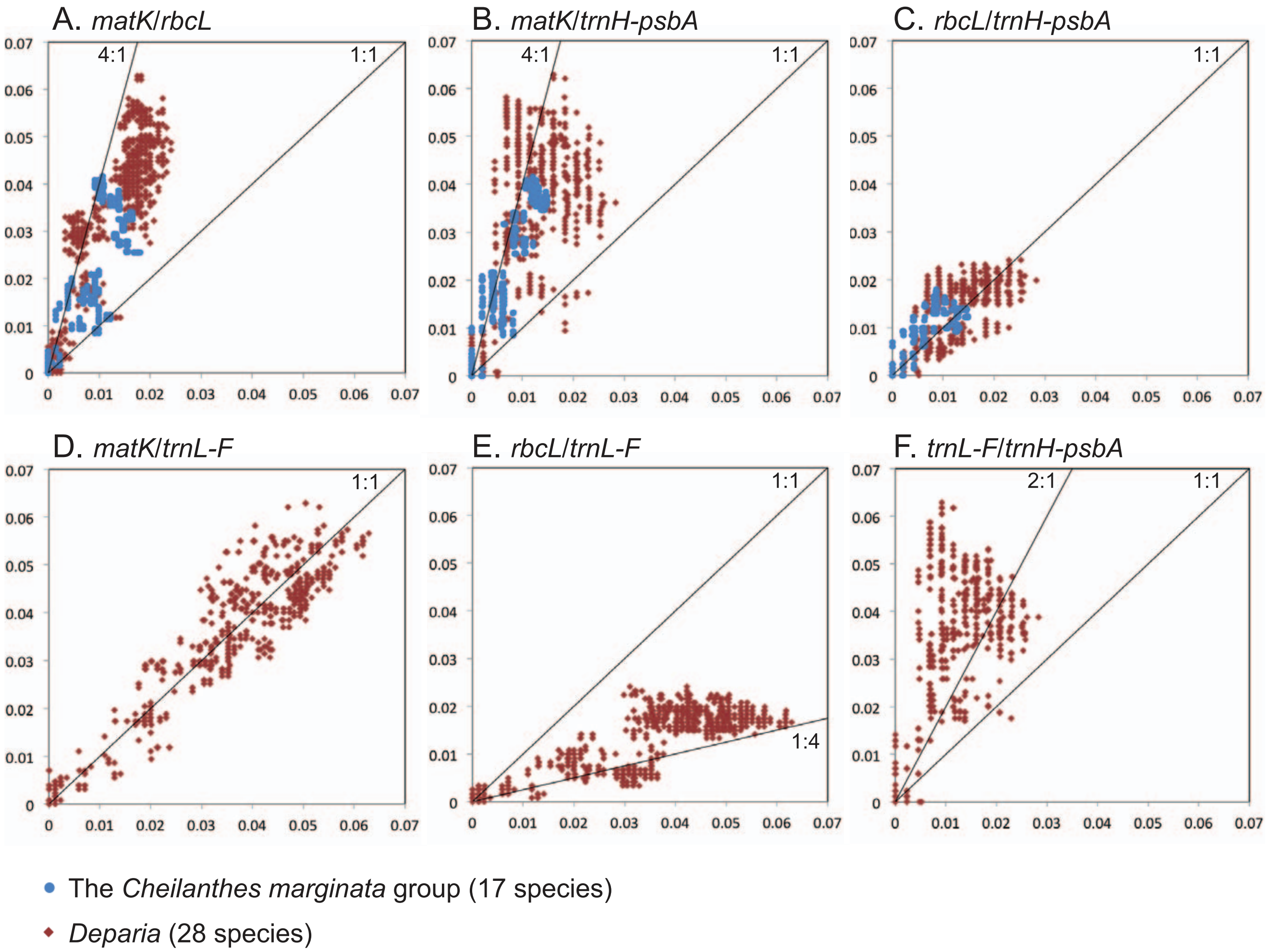

Figure 2. Loci comparisons within the focal polypod genera: the *Cheilanthes marginata* group and *Deparia*. The x and y axes depict the interspecific pairwise *p*-distances; the *C. marginata* group data points are represented by blue circles and *Deparia* by red diamonds. Note that *trnL-F* data were not available for the *C. marginata* group, hence they are not shown in panels D, E, F. All loci comparisons are presented as y-axis vs x-axis: (A) *matK* vs *rbcL*, (B) *matK* vs *trnH-psbA*, (C) *rbcL* vs *trnH-psbA*, (D) *matK* vs *trnL-F*, (E) *rbcL* vs *trnL-F*, (F) *trnL-F* vs *trnH-psbA*. The lines in each panel (labeled with the ratios 1:1, 1:4, 4:1 or 2:1) are not regression lines, but are drawn to guide the eye in interpreting the results.
doi:10.1371/journal.pone.0026597.g002

Table 2. The power of barcoding loci and locus combinations to discriminate species.

	Percentage of species that can be uniquely discriminated (%)	
	***Cheilanthes marginata* group**	***Deparia***
matK	47.1	75.0
rbcL	41.2	67.9
trnH-psbA	17.6	46.4
trnL-F	-	57.1
matK+rbcL[1]	47.1	100
matK+trnH-psbA	41.2	92.9
rbcL+trnH-psbA	41.2	82.1
matK+rbcL+trnH-psbA	47.1	100
matK+trnL-F	-	82.1
rbcL+trnL-F	-	78.6
matK+rbcL+trnL-F	-	100

[1] the official two-locus DNA barcode [18,19].
doi:10.1371/journal.pone.0026597.t002

Materials and Methods

Sampling

To assess the discriminatory power of potential DNA barcoding loci at a broad scale, we assembled sequences of *rbcL*, *matK* and *trnH-psbA* from 74 fern species, including representatives from each of the 37 families (*sensu* [29]), and *trnL-F* sequences from 32 species representing 19 families (Supporting Information Table S2 and Table S4). More intensive inter- and intra-specific sampling was done within two focal polypod genera: *Deparia* and the *Cheilanthes marginata* group. In a recent phylogenetic analysis of the cheilanthoid ferns (M.D. Windham et al., unpublished), the *C. marginata* group was shown to be strongly monophyletic and only distantly related to the type species of *Cheilanthes* (*C. micropteris* Swartz). Taxonomic treatment of this group as a new genus is forthcoming (F.W. Li et al., unpublished). This new genus comprises 17 species, all of which were included in this study. Fifteen of the 17 species were represented by multiple individuals, giving a total sample size of 58 for the *C. marginata* group (Supporting Information Table S2). The two species that lack intra-specific sampling are known only from the type specimen or a single non-type collection (see Supporting Information Table S2) [47]. All members of the *C. marginata* group were sequenced for *matK*, *rbcL* and *trnH-psbA*. The genus *Deparia* comprises approximately 50 species currently assigned to the Woodsiaceae in the eupolypod II lineage [29,51,52], and 28 of these were sampled for this study. Three *Deparia* species included intra-specific sampling (42 individuals in total; Supporting Information Table S2 and Table S4), and two varieties were treated as distinct species for the species discrimination tests (Table 2). Four loci, *matK*, *rbcL*, *trnH-psbA*, and *trnL-F* were sequenced in *Deparia*. Voucher information for all taxa included in this study are provided in the Supporting Information Table S2 and Table S4, along with their GenBank sequence accessions.

DNA Extraction, Amplification and Sequencing

For the *Cheilanthes marginata* group, genomic DNA was extracted using QIAGEN DNeasy Plant Mini Kits following published protocols [51]. DNA extractions in *Deparia* were done using a modified CTAB procedure [53]. All primers and PCR conditions used in this study are reported in Supporting Information Table S3. For the large-scale locus comparisons, 17 taxa were newly sequenced for *trnH-psbA* using "trnH2" and "psbAF" primers [54] or specially designed universal primers. In the *C. marginata* group, *trnH-psbA* amplification and sequencing were mostly done using "trnH2" and "psbAF" [54]; in *Deparia* newly designed universal primers were used. For *matK*, specific primers were designed separately for the two focal genera. Amplification and sequencing of *rbcL* for the *C. marginata* group used published primers [51], and for *Deparia* used "F1F" [55] and "1379R" [56]. *trnL-F* was amplified and sequenced using "f" [57] and "FernLr1". For some of the older herbarium-derived samples in the *C. marginata* group, smaller overlapping fragments of *matK* and *rbcL* were amplified and sequenced using newly designed primers, and the final sequences assembled from contigs.

Sequence Alignment And Barcoding Utility Assessment

For the large-scale comparison, we calculated species pairwise sequence divergence values within each fern order [29] to minimize alignment ambiguities: Equisetales, Marattiales, Ophioglossales+Psilotales, Osmundales, Hymenophyllales, Gleicheniales, Schizaeales, Salviniales, Cyatheales, and Polypodiales. *Psilotum nudum* was our sole representative of Psilotales, and hence was compared with Ophioglossales, which belongs to the same class (Psilotopsida). Sequences in each order were separately aligned, manually for *rbcL* and *matK* and using SATé 1.2 [58] or ClustalW [59] (followed by manual adjustments) for *trnH-psbA* and *trnL-F*. In SATé, MAFFT [60] was used as the "Aligner", OPAL [61] as the "Merger", and RAxML [62] with GTRGAMMAI as the tree estimator, and other parameters followed the default settings. When there were less than four sequences to be aligned, SATé was not applicable, and ClustalW was used instead (with the default settings). Alignments within the two focal genera were straightforward and done manually.

PAUP* v4.0a114 [63] was used to calculate pairwise sequence divergence (uncorrected *p*-distance). Substitutions in sites with gaps and/or missing data were distributed proportionally to unambiguous changes in PAUP*. For comparing sequence variation of different loci, Wilcoxon matched-pairs signed-rank tests were carried out using an online calculator (http://www.fon.hum.uva.nl/Service/Statistics/Signed_Rank_Test.html). To assess the discrimination power of each DNA region, we examined the ability of each locus and locus combination to uniquely discriminate a species from all others. The success rate of species discrimination is the percentage of species that could be distinguished among all possible species pairs. A pair of species was scored as successfully distinguished if the interspecific distance was always greater than zero and greater than the intraspecific distance. A Perl script was written to calculate the discrimination success rate from the PAUP* output (available upon request).

Supporting Information

Table S1 List of species in the *Cheilanthes marginata* group that cannot be uniquely discriminated by the core DNA barcode.
(XLS)

Table S2 List of the taxa, samples, and GenBank accession used in this study, with separate sub-tables for the large-scale, *Cheilanthes marginata* group, and *Deparia* datasets.
(XLS)

Table S3 List of the primers and PCR conditions used in this study.
(XLS)

Table S4 List of the references cited in the three supporting tables.
(DOC)

Acknowledgments

We thank Dr. Chun-Neng Wang for his initial support on the fern *matK* project. We are grateful for the kind help and permission to sample herbarium specimens from the curators and staff of NY, TEX/LL, CAS and MO; for valuable comments from Dr. Layne Huiet, Erin Sigel, Amanda Grusz, and two anonymous reviewers; to Joel Nitta for sharing some unpublished information; and to Dr. George Yatskievych for supporting FWL's presentation of these results at the International Botanical Congress 2011.

Author Contributions

Conceived and designed the experiments: FWL LYK MDW KMP. Performed the experiments: FWL LYK. Analyzed the data: FWL. Contributed reagents/materials/analysis tools: FWL LYK CJR AE WLC MDW KMP. Wrote the paper: FWL MDW KMP.

References

1. Pryer KM, Schuettpelz E, Wolf PG, Schneider H, Smith AR, et al. (2004) Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. Am J Bot 91: 1582–1598.
2. Ebihara A, Farrar DR, Ito M (2008) The sporophyte-less filmy fern of Eastern North America *Trichomanes Intricatum* (Hymenophyllaceae) has the chloroplast genome of an Asian species. Am J Bot 95: 1645–1651.
3. Dassler CL, Farrar DR (2001) Significance of gametophyte form in long-distance colonization by tropical, epiphytic ferns. Brittonia 53: 352–369.
4. Rumsey FJ, Jermy AC, Sheffield E (1998) The independent gametophytic stage of *Trichomanes speciosum* Willd. (Hymenophyllaceae), the Killarney Fern and its distribution in the British Isles. Watsonia 22: 1–19.
5. Li FW, Tan BC, Buchbender V, Moran RC, Rouhan G, et al. (2009) Identifying a mysterious aquatic fern gametophyte. Pl Syst Evol 281: 77–86.
6. Farrar DR (1967) Gametophytes of four tropical fern genera reproducing independently of their sporophytes in the southern appalachians. Science 155: 1266–1267.
7. Ebihara A, Matsumoto S, Ito M (2009) Hybridization involving independent gametophytes in the *Vandenboschia radicans* complex (Hymenophyllaceae): a new perspective on the distribution of fern hybrids. Mol Ecol 18: 4904–4911.
8. Gureyeva II (2003) Demographic studies of homosporous fern populations in south Siberia. In: Chandra S, Srivastava M, eds. Pteridology in the new millennium. Dordrecht: Kluwer Academic Publishers. pp 341–364.
9. Watkins JE, Mack MK, Mulkey SS (2007) Gametophyte ecology and demography of epiphytic and terrestrial tropical ferns. Am J Bot 94: 701–708.
10. Schneider H, Schuettpelz E (2006) Identifying fern gametophytes using DNA sequences. Mol Ecol Notes 6: 989–991.
11. de Groot GA, During HJ, Maas JW, Schneider H, Vogel JC, et al. (2011) Use of *rbcL* and *trnL-F* as a two-locus DNA barcode for identification of NW-European ferns: an ecological perspective. PLoS ONE 6: e16371.
12. Li FW, Kuo LY, Huang YM, Chiou WL, Wang CN (2010) Tissue-direct PCR, a rapid and extraction-free method for barcoding of ferns. Mol Ecol Resour 10: 92–95.
13. Pryer KM, Schuettpelz E, Huiet L, Grusz AL, Rothfels CJ, et al. (2010) DNA barcoding exposes a case of mistaken identity in the fern horticultural trade. Mol Ecol Resour 10: 979–985.
14. Zhao ZL, Leng CH, Wang ZT (2007) Identification of *Dryopteris crassirhizoma* and the adulterant species based on cpDNA *rbcL* and translated amino acid sequences. Planta Med 73: 1230–1233.
15. Ma XY, Xie CX, Liu C, Song JY, Yao H, et al. (2010) Species identification of medicinal pteridophytes by a DNA barcode marker, the chloroplast *psbA-trnH* intergenic region. Biol Pharm Bull 33: 1919–1924.
16. Lahaye R, van der Bank M, Bogarin D, Warner J, Pupulin F, et al. (2008) DNA barcoding the floras of biodiversity hotspots. Proc Nat Acad Sci USA 105: 2923–2928.
17. Hollingsworth ML, Andra Clark A, Forrest LL, Richardson J, Pennington RT, 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 Resour 9: 439–457.
18. CBOL Plant Working Group website. Available: http://www.barcoding.si.edu/plant_working_group.html. Accessed 2011 September, 30.
19. Hollingsworth PM, Graham SW, Little DP (2011) Choosing and using a plant DNA barcode. PLoS ONE 6: e19254.
20. 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.
21. Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, et al. (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE 3: e2802.
22. CBOL Plant Working Group (2009) A DNA barcode for land plants. Proc Nat Acad Sci USA 106: 12794–12797.
23. 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.
24. Johnson LA, Soltis DE (1995) Phylogenetic inference in Saxifragaceae *sensu stricto* and *Gilia* (Polemoniaceae) using *matK* sequences. Ann Mo Bot Gard 82: 149–175.
25. Hilu KW, Borsch T, Muller K, Soltis DE, Soltis PS, et al. (2003) Angiosperm phylogeny based on *matK* sequence information. Am J Bot 90: 1758–1776.
26. Duffy AM, Kelchner SA, Wolf PG (2009) Conservation of selection on *matK* following an ancient loss of its flanking intron. Gene 438: 17–25.
27. Wolf PG, Rowe CA, Sinclair RB, Hasebe M (2003) Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. DNA Res 10: 59–65.
28. Kuo LY, Li FW, Chiou WL, Wang CN (2011) First insights into fern *matK* phylogeny. Mol Phylogenet Evol 59: 556–566.
29. Smith AR, Pryer KM, Schuettpelz E, Korall P, Schneider H, et al. (2006) A classification for extant ferns. Taxon 55: 705–731.
30. Kress W, Wurdack K, Zimmer E, Weigt L (2005) Use of DNA barcodes to identify flowering plants. Proc Nat Acad Sci USA 102: 8369–8374.
31. Roy S, Tyagi A, Shukla V, Kumar A, Singh UM, et al. (2010) Universal plant DNA barcode loci may not work in complex groups: a case study with Indian berberis species. PLoS ONE 5: e13674.
32. Ren B-Q, Xiang X-G, Chen Z-D (2009) Species identification of *Alnus* (Betulaceae) using nrDNA and cpDNA genetic markers. Mol Ecol Resour 10: 594–605.
33. Newmaster SG, Ragupathy S (2009) Testing plant barcoding in a sister species complex of pantropical *Acacia* (Mimosoideae, Fabaceae). Mol Ecol Resour 9: 172–180.
34. Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J (2008) Testing candidate plant barcode regions in the Myristicaceae. Mol Ecol Resour 8: 480–490.
35. Piredda R, Simeone MC, Attimonelli M, Bellarosa R, Schirone B (2010) Prospects of barcoding the Italian wild dendroflora: oaks reveal severe limitations to tracking species identity. Mol Ecol Resour 11: 72–83.
36. Liu Y, Yan HF, Cao T, Ge XJ (2010) Evaluation of 10 plant barcodes in Bryophyta (Mosses). J Syst Evol 48: 36–46.
37. Nitta JH (2008) Exploring the utility of three plastid loci for biocoding the filmy ferns (Hymenophyllaceae) of Moorea. Taxon 57: 725–736.
38. Wolf PG, Roper JM, Duffy AM (2010) The evolution of chloroplast genome structure in ferns. Genome 53: 731–738.
39. Curtis SE, Clegg MT (1984) Molecular evolution of chloroplast DNA sequences. Mol Biol Evol 1: 291–301.
40. Bowman CM, Bonnard G, Dyer TA (1983) Chloroplast DNA variation between species of *Triticum* and *Aegilops* — location of the variation on the chloroplast genome and its relevance to the inheritance and classification of the cytoplasm. Theor Appl Genet 65: 247–262.
41. Palmer JD, Singh GP, Pillay DTN (1983) Structure and sequence evolution of three legume chloroplast DNAs. Mol Gen Genet 190: 13–19.
42. Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Nat Acad Sci USA 84: 9054–9058.
43. Birky CW, Walsh JB (1992) Biased gene conversion, copy number, and apparent mutation rate differences within chloroplast and bacterial genomes. Genetics 130: 677–683.
44. Perry AS, Wolfe KH (2002) Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. J Mol Evol 55: 501–508.
45. Spooner DM (2009) DNA barcoding will frequently fail in complicated groups: an example in wild potatoes. Am J Bot 96: 1177–1189.
46. Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, Graham SW, et al. (2009) Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? Mol Ecol Resour 9 (Suppl. 1): 130–139.
47. Mickel JT, Smith AR (2004) The pteridophytes of Mexico. Memoir New York Bot Gard 88: 176–213.
48. Yao H, Song JY, Liu C, Luo K, Han JP, et al. (2010) Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS ONE 5: e13102.
49. Chen S, Yao H, Han J, Liu C, Song J, et al. (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS ONE 5: e8613.
50. Van den Heede CJ, Viane RLL, Chase MW (2003) Phylogenetic analysis of *Asplenium* subgenus *Ceterach* (Pteridophyta: Aspleniaceae) based on plastid and nuclear ribosomal ITS DNA sequences. Am J Bot 90: 481–495.
51. Schuettpelz E, Pryer KM (2007) Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. Taxon 56: 1037–1050.
52. Rothfels CJ, Larsson A, Kuo LY, Korall P, Chiou WL, et al. In press.
53. Wang CN, Moller M, Cronk QC (2004) Phylogenetic position of *Titanotrichum oldhamii* (Gesneriaceae) inferred from four different gene regions. Syst Bot 29: 407–418.
54. Tate JA, Simpson BB (2003) Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. Syst Bot 28: 723–737.
55. Wolf PG, Soltis PS, Soltis DE (1994) Phylogenetic relationships of dennstaedtioid ferns: evidence from *rbcL* sequences. Mol Phylogenet Evol 3: 383–392.
56. Pryer KM, Smith AR, Hunt JS, Dubuisson JY (2001) *rbcL* data reveal two monophyletic groups of filmy ferns (Filicopsida: Hymenophyllaceae). Am J Bot 88: 1118–1130.
57. Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three noncoding regions of chloroplast DNA. Plant Mol Biol 17: 1105–1109.
58. Liu K, Raghavan S, Nelesen S, Linder CR, Warnow T (2009) Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. Science 324: 1561–1564.
59. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680.
60. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform 9: 286–298.
61. Wheeler TJ, Kececioglu JD (2007) Multiple alignment by aligning alignments. Bioinformatics 23: i559–568.
62. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690.
63. Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.