Exploring the utility of three plastid loci for biocoding the filmy ferns (Hymenophyllaceae) of Moorea

Joel H. Nitta

University of California at Berkeley, Department of Integrative Biology, 4156 Valley Life Sciences Building, Berkeley, California 94720, U.S.A. jnitta@berkeley.edu. Current address: Department of Biological Sciences, Graduate School of Science, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

The Moorea Biocode seeks to compile an integrative taxonomy based on morphological, ecological, and molecular data for the entire macrobiota of the island of Moorea, French Polynesia. As a case study exploring the utility of selected molecular data for species identification within this integrative taxonomic framework, chloroplast DNA from three regions (*rbcL*, *trnSGG*, *trnH-psbA*) were sequenced for all species of filmy ferns (Hymenophyllaceae) known from Moorea. The relative utility of each of these regions for the purposes of phylogenetic analysis and DNA-based identification was inferred by estimating support for phylogenetic trees reconstructed from each region and by calculating intraspecific and interspecific distance values (uncorrected *p*) between taxa for each region. All three of these regions were found to be potentially useful for phylogenetic studies at the appropriate taxonomic level. *trnH-psbA* was determined to have the greatest utility as a potential marker for DNA-based identification because of its high interspecific variability and high degree of amplification success. *rbcL* and *trnH-psbA* were successfully used in combination with morphological characters to identify a previously unidentified Moorean filmy fern species, *Polyphlebium borbonicum*. The presented results support *trnH-psbA* as appropriate for completing the plant section of the Moorea Biocode, although the use of additional markers will be necessary.

KEYWORDS: DNA barcoding, French Polynesia, Hymenophyllaceae, integrative taxonomy, Moorea, Moorea Biocode, plastid DNA, *rbcL*, *trnSGG*, *trnH-psbA*

INTRODUCTION

Taxonomic identification based on molecular data is neither new nor controversial; this technique has been practiced ad hoc since the technology to obtain molecular data became widely available (e.g., Fox & al., 1977; Baker & Palumbi, 1994; DeSalle & Birstein, 1996; Amato & al., 1999; Schneider & Schuettpelz, 2006). A recent formulation of this technique, under the name "DNA barcoding" (Hebert & al., 2003a), seeks to standardize the process by using one or a few universal genetic markers (i.e., barcodes) for species identification across all life. A portion of the mitochondrial cyctochrome oxidase I gene (COI or *coxI*) has been largely accepted as the standard barcode marker in animals and its efficacy has been supported by a number of case studies (Hebert & al., 2003b, c; Barrett & Hebert, 2005; Ward & al., 2005; Hajibabaei & al., 2006); however, other studies demonstrate problems with a single-gene COI approach (Meyer & Paulay, 2005; Prendini, 2005; Brower, 2006; Meier & al., 2006), and the biological and phylogenetic validity of DNA barcoding has been the topic of a great deal of debate (Libscomb & al., 2003; Blaxter, 2004; Will & Rubinoff, 2004; Smith, 2005).

Attempts to use COI or other mitochondrial loci as barcode markers in plants has failed due to the low rate of mitochondrial genome evolution in plants compared to animals and most other organisms. The chloroplast genome has been suggested as a likely source of candidate barcode markers because of its single-copy nature, uniparental inheritance, and relative ease of isolation compared to nuclear DNA (Kress & al., 2005). Although it is unlikely that any single chloroplast locus exists with interspecific sequence divergence rates as high as COI in animals (Chase & al., 2007), several regions including both protein-coding genes (*rbcL*, *matK*, *rpoC1*, *rpoB*) and non-coding spacers (*trnH-psbA* intergenic spacer, *trnL* intron) have been identified as possible barcode candidates (Kress & al., 2005; Taberlet & al., 2007), and there have been recent proposals to combine two or more of these regions to barcode the plants (Newmaster & al., 2006; Chase & al., 2007; Kress & Erickson, 2007). The lack of consensus for a global barcode marker in plants and the controversy concerning DNA barcoding reflect the inherent limits of this technique: DNA sequence data from one standard locus alone will never be sufficient to distinguish all species, especially in recently diverged species. These limitations must be resolved by considering not only selected DNA sequences but also other kinds of evidence as proposed by the concept of integrative taxonomy (Dayrat, 2005; Will & al., 2005; Mengual & al., 2006).

The Moorea Biocode project is a case study in integrative taxonomy currently underway on the island of Moorea, French Polynesia. The project seeks to "biocode," or collect DNA sequence data, distribution data, images, and morphological and ecological descriptions of, every macroscopic species on the island of Moorea, French Polynesia (Check, 2006). Collection information, location, and availability of voucher specimens are accessible through an online database (http://bscit.berkeley.edu/biocode). The Moorea Biocode project takes a clade-based approach, directing collection efforts towards specific monophyletic groups one at a time. This provides a phylogenetic framework so that as new vouchers are added to the database, they can be analyzed in terms of closely related specimens. Now that several Moorean animal lineages including fish and spiders have been added to the database, candidate plant groups are being considered for sampling.

Ferns (monilophytes) and lycophytes have received high priority for the Moorea Biocode Project. Both lineages contribute important components to insular plant diversity and are overrepresented relative to angiosperms on islands compared to continents, a fact most often attributed to their putative high dispersal ability (Fosberg, 1948; Smith, 1972; Dassler & Farrar, 2001). Of the plant groups present on Moorea, the ferns and lycophytes are the best documented (Gulamhussein, 2000; Murdock & Smith, 2003; Ranker & al., 2005; Nitta, 2006), and there are many available recent collections due to ongoing work on the French Polynesian flora by J. Florence and collaborators.

As a case study on the utility of integrative taxonomic methods, and as a first step toward cataloging the ferns of Moorea within this framework, the filmy ferns (Hymenophyllaceae—see Fig. 1) were chosen as the most suitable study group. Filmy ferns have been shown to constitute a major part of the fern flora on many oceanic islands (Dassler & Farrar, 2001), and Moorea is no exception (12 of 82 species, or 14.6% of the total fern flora; Murdock & Smith, 2003; Ranker & al., 2005; Nitta, 2006). The filmy ferns of Moorea are a diverse group, representing six of nine genera sensu Ebihara & al. 2006. Thus, they serve as an excellent snapshot of the morphological and ecological diversity contained within the family as a whole. The filmy ferns of Moorea have been sampled intermittently over several fern collection efforts since the 1930s (Murdock & Smith, 2003). Gulamhussein (2000) conducted an in-depth ecological study of three species of *Trichomanes* s.l. on Moorea. The morphology, distribution, and ecology of all twelve known Moorean filmy ferns were surveyed by Nitta (2006), leaving only molecular data wanting to complete the integrative taxonomy of the group.

Because of recent molecular phylogenetic studies on filmy ferns, significant amounts of sequence data are publicly available; however, it is unclear whether the markers currently used for phylogenetic purposes will be appropriate for species-level identification, and the need for additional phylogenetic markers in filmy ferns has also been pointed out (Hennequin & al., 2006). The majority of barcoding efforts in plants have focused on angiosperms, and the suitability of various proposed barcode markers has not been assessed for ferns. The goals of this study are: *(1)* to assess candidate chloroplast loci for use as identification and phylogenetic markers in filmy ferns, and *(2)* to explore methods of identification using multiple data types. This case study on filmy ferns can be used to determine how to best proceed in completing the integrative taxonomic work for the rest of the plants of Moorea, and methods developed here can be applied to similar integrative taxonomic studies in the future.

MATERIALS AND METHODS

Marker selection. — Three chloroplast regions were selected to determine their suitability as identification and phylogenetic markers in the Moorean filmy ferns: *rbcL*, the *trnH*^{GUG}*-psbA* intergenic spacer (hereafter referred to as "*trnH-psbA*"), and the *trnS*^{GCU}*-trnG*^{UUC} intergenic spacer + *trnG*^{UUC} intron (hereafter referred to as "*trnSGG*"). *rbcL* (ca. 1,240 bp) and *trnH-psbA* (ca. 250 to 500 bp) are two of the leading candidate plant barcode markers based on results of studies focusing primarily on flowering plants (Kress & al., 2005; Cowan & al., 2006; Kress & Erickson, 2007), and their utility needs to be verified in other groups such as ferns. *trnSGG* (ca. 1,700 bp) is composed primarily of non-coding spacer DNA and has been suggested as a source of phylogenetic characters in ferns (Small & al., 2005); it also contains two short coding regions, *psaM* and *ycf12* (ca. 100 bp each). *trnSGG* has never been tested for use as an identification marker in any group.

Sampling. — The study taxa included all twelve filmy fern species known to occur on Moorea. The selected chloroplast regions were sequenced for two to three specimens per species, resulting in 72 Moorean sequences from 27 specimens total. In order to better characterize within-species variation, several specimens from outside of Moorea were also sequenced for the selected chloroplast regions (24 "additional location" sequences from 10 specimens total); these were either collected in the field or came from herbarium specimens at the University of California, Berkeley (UC). In some cases, one or more regions could not be successfully amplified for a specimen; for amplification success rates, see Table 1. Additionally, all available Hymenophyllaceae *rbcL* data on GenBank (http://www.ncbi.nlm.nih.gov/Genbank/) alignable within a 1,206 bp region were obtained in order to aid in species identification and to characterize variation in this gene

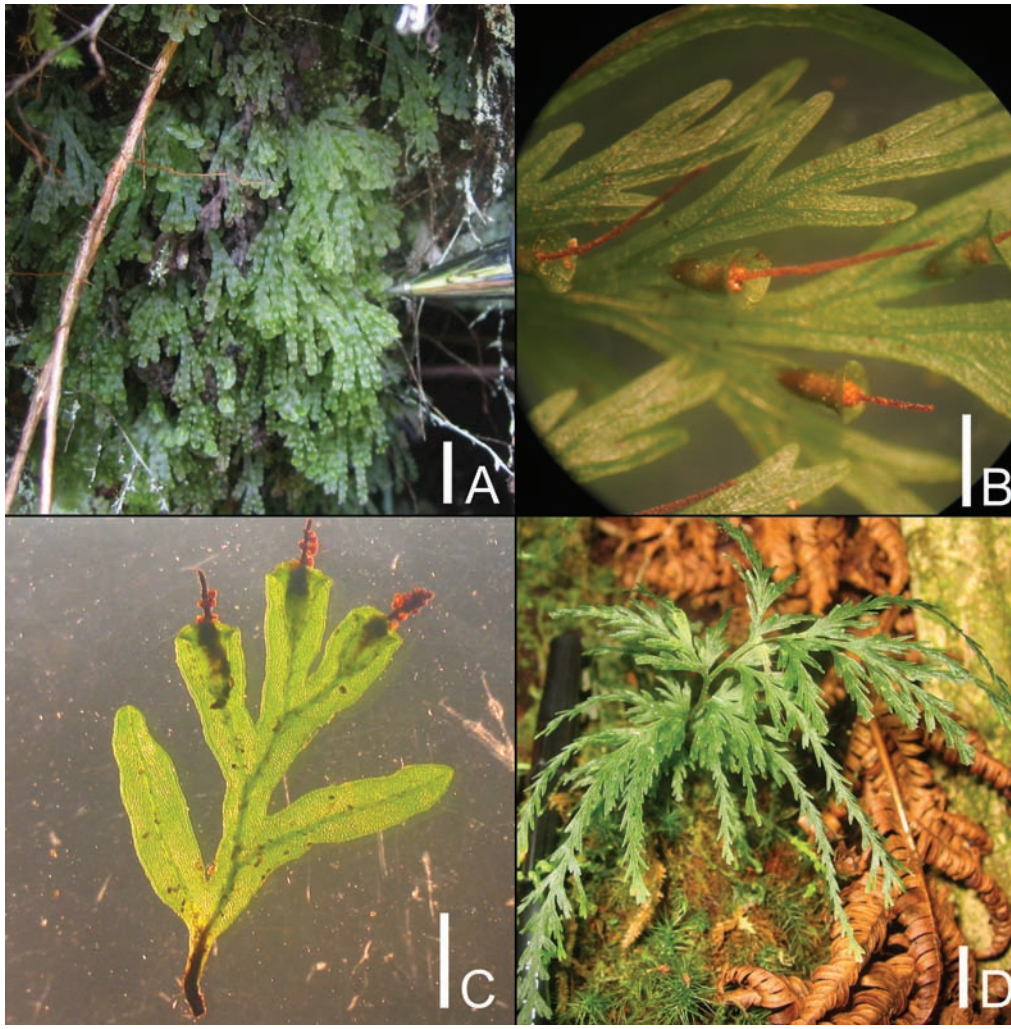

Fig. 1. Examples of Moorean filmy ferns. A, colony of *Hymenophyllum digitatum* (Sw.) Fosberg (specimen *JN025*) growing at base of *Metrosideros collina* (J.R. & G. Forst.) A. Gray, scale bar = 1 cm; B, involucres and sori of *Abrodictyum caudatum* (Brack.) Ebihara & K. Iwats. (specimen *JN031*), scale bar = 1 mm; C, fertile *Crepidomanes kurzii* (Bedd.) Tagawa & K. Iwats. frond (specimen *JN018*), scale bar = 1 mm; D, *Polyphlebium borbonicum* (Bosch) Ebihara & Dubuisson (specimen *JN073*) found on Mt. Mouaputa, scale bar = 1 cm. All photos by Joel Nitta.

Table 1. DNA sequence and amplification success statistics by chloroplast region.

	*rbcL**	*rbcL*	*trnSGG*	*trnH-psbA*	Combined
Aligned sequence length	1,206	1,238	2,334	782	4,354
# (%) variable characters	538 (44.6%)	297 (24.0%)	1,283 (55.0%)	562 (71.8%)	2,142 (49.2%)
# (%) pars. inform. chars.	449 (37.2%)	288 (23.3%)	1,108 (47.5%)	492 (62.9%)	1,888 (43.4%)
Max. sequence length	–	–	1,782	565	–
Min. sequence length	–	–	1,556	218	–
Mean sequence length	–	–	1,718 ± 62	426 ± 92	–
# sequences	–	35	28	33	–
% success	–	95%	76%	89%	–

* indicates broad sampling (GenBank, Moorean, and "additional locality" *rbcL*), other chloroplast regions are from restricted sampling (Moorean and "additional locality" sequences only). *trnSGG* and *trnH-psbA* datasets include binary gap characters (see Results). # sequences and % success indicate amplification success out of 37 specimens attempted. *rbcL* is not length variable; all sequences were 1,238 bp in length after editing for ambiguous bases. – indicates non-applicability.

across the family (154 species in 8 genera, 202 GenBank sequences total). For a summary of all specimens included in this study see Appendix 1 (newly sequenced species) and Appendix 2 (*rbcL* data: see Taxon online issue).

DNA extraction, amplification, and sequencing. — Prior to extraction, samples were carefully examined under a dissecting microscope and any potential contaminants (e.g., epiphyllous bryophytes and algae) removed. Total DNA extraction was performed with the Plant DNeasy Mini Kit (Qiagen) following the manufacturer's protocol. PCR amplification was performed in 20 μL reactions containing AccuPower PCR PreMix (Bioneer), 10 μM forward primer, 10 μM reverse primer, 2 μL of total genomic DNA, and de-ionized water to volume. *rbcL* was amplified using primers and PCR protocol outlined in Pryer & al. (2001). *trnH-psbA* was amplified using the primers of Tate & Simpson (2003) following the same PCR protocol for *rbcL*. *trnSGG* was amplified using primers and PCR protocol as described by Small & al. (2005). PCR products were checked for successful amplification on a 1% agarose gel in TAE buffer. Excess primers and nucleotides were removed using ExoSapIT (Amersham Pharmacia) according to the manufacturer's protocol. Automated sequencing of purified PCR products was performed on a 3730xl DNA Analyzer (Applied Biosystems) at the University of California, Berkeley Sequencing Center. Internal primers trnG5′2G and trnG5′2S (Shaw & al., 2005) were also used for sequencing of the *trnSGG* region.

Sequence alignment. — DNA sequences were assembled using Sequencher 4.1 (Gene Codes Corporation). Alignment of *rbcL* was trivial and performed by eye using MacClade 4.0 (Maddison & Maddison, 2000). Sequence length varied in the noncoding regions *trnSGG* and *trnH-psbA* due to extensive insertion/deletion events. These sequences were aligned with the program MUSCLE (Edgar, 2004) on default settings and edited by eye using MacClade 4.0 to correct for obvious misalignments. Simple gap coding (Simmons & Helga, 2000) was performed using GapCoder (Young & Healy, 2003). Due to the length of *trnSGG* and the position of the internal primers used, ca. 100 bp from the middle of this region were not reliably sequenced in many cases and were removed from analysis.

Phylogenetic analysis. — Phylogenetic trees were inferred for two sample groups: broad sampling included *rbcL* data from GenBank, Moorean, and "additional location" filmy ferns (237 taxa total); restricted sampling included sequence data of all three chloroplast regions for Moorean and "additional location" specimens only (37 taxa total). Restricted sampling explored combining the chloroplast regions into several different datasets: one dataset for each region (three single-region datasets total), one dataset for each combination of two regions (three two-region datasets total), and one dataset including all three regions. Number of taxa in each dataset varied slightly due to missing data. The nearest extant sister family to Hymenophyllaceae (i.e., Gleicheniaceae; Pryer & al., 2004) is too distant to be reliably included because of difficulty in alignment. Therefore, an outgroup was not specified and trees were rooted a posteriori between the two major clades retrieved by phylogenetic analysis: the trichomanoid clade (roughly corresponding to *Trichomanes* s.l.) and the hymenophylloid clade (roughly corresponding to *Hymenophyllum* s.l.; Smith & al., 2006); the reciprocal monophyly of these clades has been confirmed in several phylogenetic studies (Pryer & al., 2001; Dubuisson & al., 2003; Pryer & al., 2004; Ebihara & al., 2006; Hennequin & al., 2006).

Parsimony analyses were conducted in PAUP* version 4.0b10 (Swofford, 2002) using a heuristic search (10,000 random addition sequences, TBR branch swapping). Support values for the maximum parsimony (MP) trees were estimated with 1,000 bootstrap replicates (Felsenstein, 1981) with ten random addition sequences per replicate using a heuristic search with the same settings as the parsimony analyses.

For Bayesian analysis of the broad sampling, all characters were included in a single partition. For the restricted sampling, data were split into seven partitions: 1) *rbcL*, 2) non-coding *trnsGG*, 3) *psaM*, 4) *ycf12*, 5) *trnSGG* gap characters, 6) *trnH-psbA*, and 7) *trnH-psbA* gap characters. The *rbcL* dataset included partition 1 only; the *trnSGG* dataset included partitions 2–5, and the *trnH-psbA* dataset included partitions 6 and 7. The two gap character partitions (5 and 7) were treated as binary (restriction) data; others were treated as DNA data. The Akaike Information Criterion as implemented in ModelTest (Posada & Crandall, 1998) was used to determine the most appropriate model of evolution to apply to each DNA data partition. For datasets that included multiple partitions, parameter values were allowed to vary between partitions. MrBayes (Huelsenbeck & Ronquist, 2001) searched treespace starting from flat priors with two sets of four chains (one cold, three hot) running simultaneously for 1,000,000 generations, sampling trees once per 100 generations. Convergence was confirmed when average standard deviation of split frequencies reached 0.01 or less. Log likelihood scores for each sample were plotted against number of generations using Tracer 1.2.1 (Rambaut & Drummon, 2003) to determine the point at which stationarity was reached; samples prior to this point were discarded as burnin (generally 25%).

Barcode utility analysis. — In order to assess utility for the purpose of DNA-based identification, average interspecific and intraspecific sequence variation (uncorrected *p*) was calculated within the broad sampling (GenBank, Moorean, and "additional location" *rbcL*), restricted

sampling (Moorean and "additional location" sequences), and Moorean sequences only using PAUP*. The resulting data were compared in three ways: by chloroplast region, by major clade retrieved in phylogenetic analysis, and by species (intraspecific variation only).

Data availability. — Voucher specimens were deposited at UC. All known morphological data, distribution information, images, and ecological data were uploaded to the Moorea Biocode online database (http://bscit.berkeley.edu/biocode). DNA data capabilities are not yet available for the Moorea Biocode database. Sequences are available on GenBank (http://www.ncbi.nlm.nih.gov/) and alignments are available through TreeBase (http://www.treebase.org).

RESULTS

DNA sequences and alignments. — Table 1 summarizes DNA sequence, alignment, and amplification success statistics. The combined dataset of the restricted sampling (72 Moorean and 24 "additional location" sequences) had a total of 4,354 characters (1,238 *rbcL* characters, 2,334 *trnSGG* characters, and 782 *trnH-psbA* characters). The *trnSGG* and *trnH-psbA* regions included 215 and 115 parsimony-informative gap regions respectively, which were coded as binary characters and included in the matrix. The combined dataset of the restricted sampling included a total of 2,142 (49.2%) variable characters and 1,888 (43.4%) parsimony-informative characters. The broad sampling alignment (202 GenBank, 27 Moorean, and 8 "additional location" *rbcL* sequences) had a final length of 1,206 bp, including 538 (44.6%) variable characters and 449 (37.2%) parsimony-informative characters.

Phylogenetic analysis. — The heuristic search of the combined dataset of the restricted sampling in PAUP* found 1,351 equally parsimonious trees of 3,972 steps each (consistency index = 0.737, retention index = 0.900). The following models were found under the Akaike Information Criterion as the most appropriate to apply to each data partition: GTR + I + Γ for *rbcL*, TVM + I + Γ for non-coding *trnSGG*, TVM + Γ for *psaM*, HKY + Γ for *ycf12*, and K81uf + I + Γ for *trnH-psbA*. Trees resulting from Bayesian analyses of the combined dataset and each single chloroplast region of the restricted sampling are shown in Figs. 2 and 3 respectively; tree topologies resulting from MP analyses did not differ except for lack of support at some nodes and are not shown, but support values for each analysis are indicated. Relationships between *Hymenophyllum* species varied between each dataset and are unresolved in the combined tree. Topology of trees resulting from analyses of combinations of two gene regions did not vary significantly from that of the combined dataset and are not shown. Two clades receiving high support values in all analyses and containing a relatively large number of taxa (trichomanoid and hymenophylloid, see Fig. 2) were chosen for analysis of barcode utility by clade.

The heuristic search of the broad sampling in PAUP* could not be completed due to insufficient processing power to analyze the dataset (237 taxa with 1,206 characters including many highly similar sequences). The GTR + I + Γ model was chosen by the Akaike Information Criterion as the most appropriate model of evolution to apply to the broad sampling dataset. An additional 4,000,000 generations (5,000,000 total) were required during Bayesian analysis for runs to converge. Since the primary goal of phylogenetic analysis of the broad sampling was to confirm support for clades for use in barcode utility analysis and not to reconstruct fine-scale topology within the family, broad sampling trees are not shown. The two clades used in barcode utility analysis were well supported by Bayesian analysis of the broad sampling (posterior probability 1.00 each). Nine out of forty-three species including more than one specimen per species were resolved as paraphyletic or polyphyletic by Bayesian analysis.

Barcode utility analysis. — Figure 4 summarizes observed rates of intraspecific and interspecific variation for broad sampling (GenBank, Moorean, and "additional locality" *rbcL* sequences) and restricted sampling (Moorean and "additional locality" sequences of all three regions) by chloroplast region. Moorean sequences alone showed no significant difference in interspecific variation from the restricted sampling and are not shown. *trnH-psbA* showed the highest mean interspecific variation (uncorrected *p*-value; 19.00% ± 5.72%, n = 89 species pairs) whereas *rbcL* showed the lowest (2.81% ± 1.67%, n = 6,708 species pairs; broad sampling). Figure 5 compares rates of variation between the two major clades retrieved in phylogenetic analysis; rates of interspecific variation were lower for the hymenophylloid clade than the trichomanoid clade in all three regions. Table 2 summarizes observed rates of intraspecific variation by species for broad and restricted sampling. Rates of intraspecific variation in Moorean sequences were extremely low for all three chloroplast regions in each species (zero for *rbcL* and *trnSGG*; 1% or less for *trnH-psbA*), with the following exceptions: 0.89% variation in *rbcL* within *Hymenophyllum pallidum* (Blume) Ebihara & K. Iwats. and 7.00% variation in *trnH-psbA* within *Crepdiomanes humile* (G. Forst.) Bosch.

DISCUSSION

Choosing a marker. — Several barcode workers have pointed out that phylogenetic and barcode markers differ in their ideal characteristics, and that markers that perform well in phylogenetic studies may not be suited for DNA-based species identification and vice-versa (Kress &

al., 2005; Newmaster & al., 2006). Given the possibly conflicting characteristics of these regions, high-throughput projects such as the Moorea Biocode Project must resolve a difficult question: which region is best to target when collecting DNA sequence data? The economical, pragmatic approach might be to choose one region that does not possess any of the ideal characteristics of either but is moderately satisfactory for both; however, one could argue that this would merely produce bad results for both phylogenetic analysis and DNA-based identification. It is probably better to sequence several regions chosen to match these different goals, in the spirit of integrative taxonomy. Therefore, suitability for both phylogenetic analysis and DNA-based identification of each marker is considered.

Selection pressures act against mutation in the in the first and second codon positions of the protein-coding *rbcL* region, limiting the amount of sequence variation

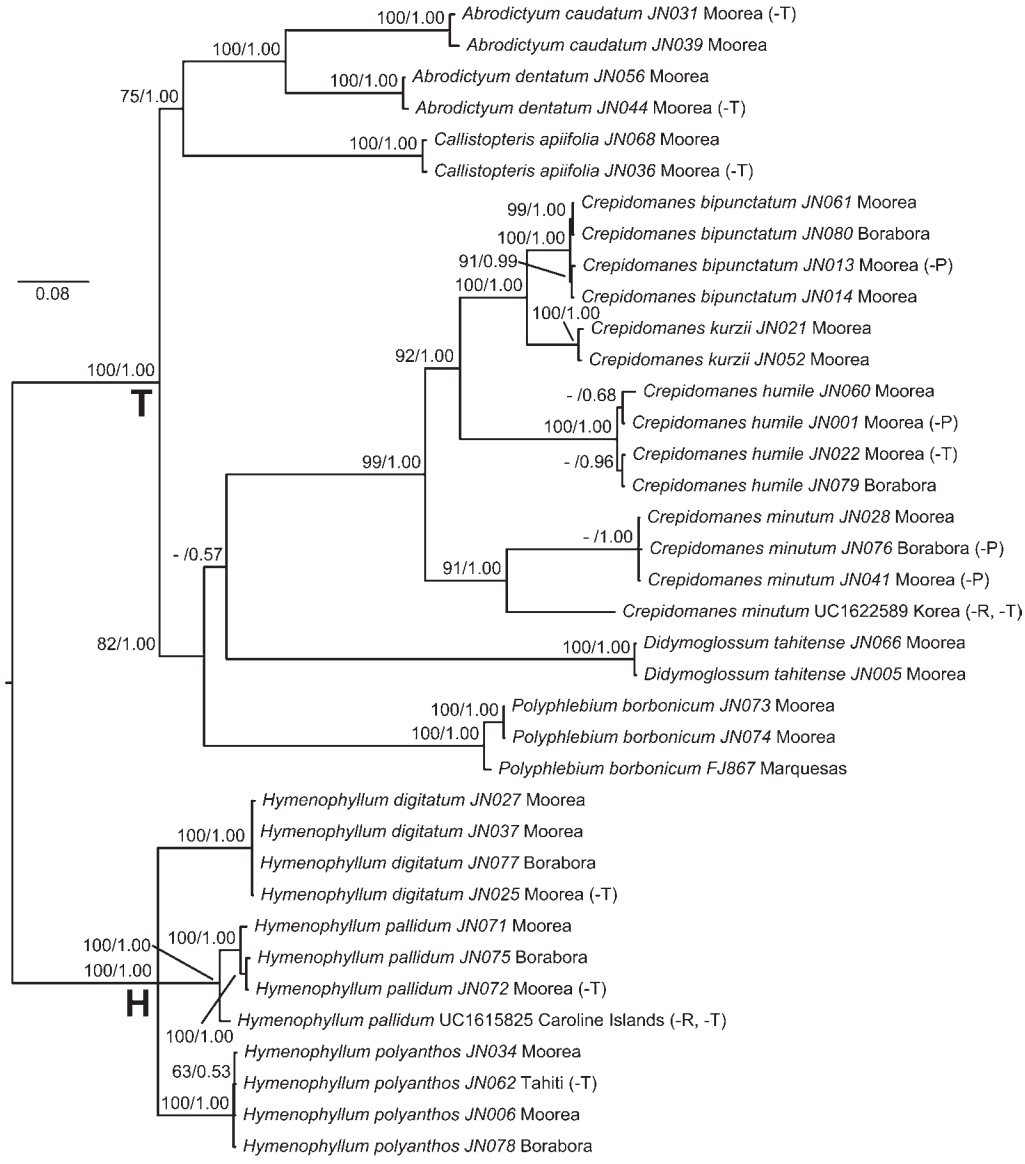

Fig. 2. Phylogeny resulting from Bayesian analysis of combined dataset (*rbcL*, *trnSGG*, and *trnH-psbA* sequences) of restricted sampling (Moorean and "additional locality" specimens). Species name, collection/herbarium number, and location indicated for each specimen. Letters in parenthesis indicate missing sequence data; R = *rbcL*, T = *trnSGG*, P = *trnH-psbA*. Numbers above nodes indicate support values (maximum parsimony bootstrap/Bayesian posterior probability); – indicates that a node was not retrieved with bootstrap value greater than 50% by MP analysis. Bold letters below nodes indicate clades used in barcode utility analysis: T = trichomanoid clade, H = hymenophylloid clade. Branchlengths are proportionate to expected amount of change per site inferred by Bayesian analysis, indicated by scalebar. Tree rooted a posterori between trichomanoid clade and hymenophylloid clade (see Methods).

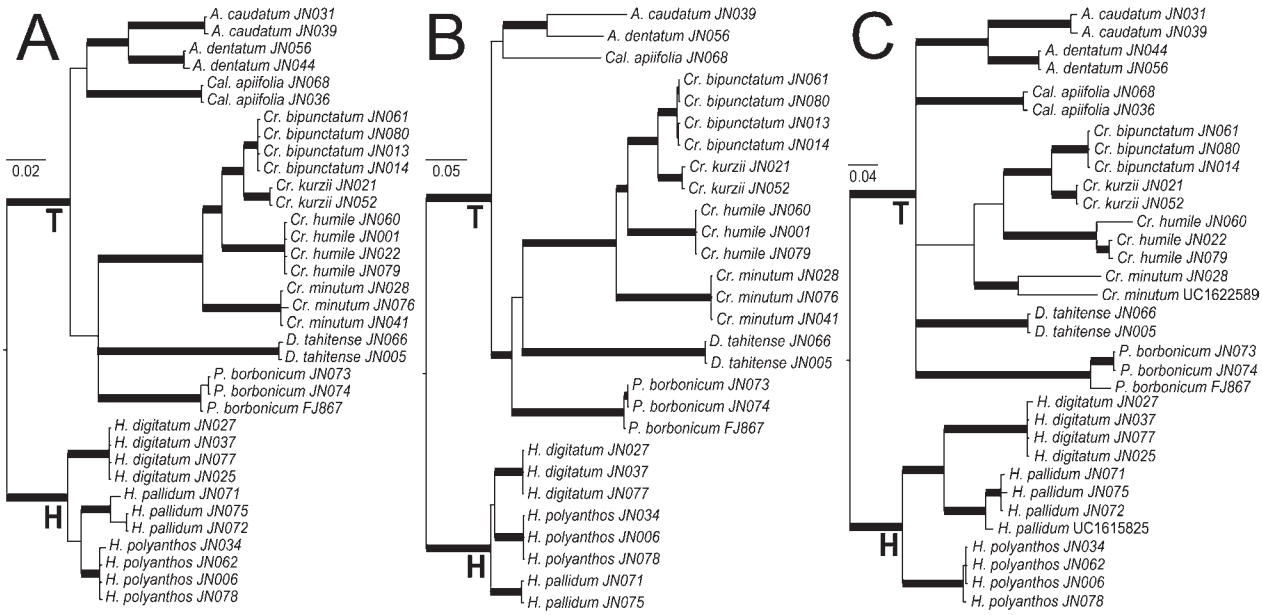

Fig. 3. Phylogenies resulting from Bayesian analysis of each chloroplast region in the restricted sampling (Moorean and "additional locality" specimens). Species name and collection/herbarium number indicated for each specimen. Bold letters below nodes indicate clades used in barcode utility analysis: T = trichomanoid clade, H = hymenophylloid clade. Branchlengths are proportionate to expected amount of change per site inferred by Bayesian analysis, indicated by scale-bar for each tree. Bold lines indicate nodes receiving good support (BS > 85, PP > 95). Trees rooted rooted *a posterori* between trichomanoid clade and hymenophylloid clade (see Methods). A, *rbcL*; B, *trnSGG*; C, *trnH-psbA*.

between taxa. This relatively low degree of variation makes *rbcL* a reliable phylogenetic marker for establishing more inclusive (i.e., basal) relationships (see internal node support with exception of *Crepidomanes*/*Didymoglossum*/*Polyphlebium* polytomy, Fig. 3A); however, this same characteristic also makes *rbcL* a poor choice as a barcode marker. A broad overlap was found at the 2%–3% divergence level in *rbcL* for interspecific and intraspecific variation (Fig. 4A); any query sequence that matches in this range cannot be reliably assigned a species identification. Additionally, one specific case of zero variation between clearly distinct species bears further discussion: *Hymenophyllum pallidum* (Blume) Ebihara & K. Iwats. and *Hymenophyllum flabellatum* Labill. These two species differ in several taxonomically important characters including involucral shape, presence of stellate hairs on lamina, and laminar thickness (*H. pallidum* is one of the few filmy ferns with laminae more than a single cell thick), and are easily differentiated in the field. However, a taxonomic identification based solely on *rbcL* data would conclude that these two ferns are the same species. This may not be an issue when the whole plant is available for morphological analysis, but it could become critical when morphological characters are unknown or cryptic, such as with gametophytes, environmental sampling, or analysis of stomach contents of herbivores. This species pair is especially relevant to Moorea, as *H. pallidum* occurs on Moorea and *H. flabellatum* is known from Tahiti. The two islands are separated by a mere 14.5 km, so it is entirely possible that *H. flabellatum* also occurs on Moorea but has not yet been observed.

Unlike *rbcL*, the *trnSGG* region consists of mostly non-coding spacer DNA, and the decreased selection pressure allows for a much higher degree of sequence variation (Fig. 4). Fewer basal nodes received strong support in *trnSGG* as opposed to *rbcL*; however, less inclusive nodes were well supported (Fig. 3B). Thus, *trnSGG* is a useful marker for lower-scale (e.g., species-level; but this is highly dependent on lineage) phylogenetic studies. This region also shows strong potential as a marker for use in DNA-based identification. Although at 2,000 bp it is considerably longer than the 800 bp maximum for barcode markers (Kress & al., 2005), primers could easily be designed within the interspersed conserved coding regions *psaM* and *ycf12* that would result in a marker under 800 bp in length. *trnSGG* displayed the high degree of divergence between species (10.83% ± 3.97%, n = 75 species pairs) and low degree of divergence within species (0.15% ± 0.15%, n = 24 conspecific pairs) required of a marker used for DNA-based identification. However, amplification success of this region was relatively low within the filmy ferns (Table 1), and primer universality must be tested in other groups to determine the ultimate practicality of using this region as a marker for DNA-based identification.

The high frequency of indel events in *trnH-psbA* (observed sequence length variation 2× or greater between

some species; see Table 1) makes confident alignment of this region difficult for closely related species, and nearly impossible for distantly related species. Consequently, support was lacking for many nodes including more than a single species in the *trnH-psbA* tree (Fig. 3C). Despite its low utility for phylogenetic analysis at the species level or higher, *trnH-psbA* is promising as a marker for DNA-based identification. This region displayed the greatest capability of the three markers to differentiate between species based on its high rate of interspecific variation; however, intraspecific variation was correspondingly high (Fig. 4), and care must be taken in determining a species-level "cutoff value" for *trnH-psbA*. Another characteristic of *trnH-psbA* that makes it attractive as a DNA-based identification marker is its ease of amplification (Table 1). This region was successfully amplified from several herbarium specimens (UC1615825, UC1622589, UC1797871) which could not be amplified for the one or both of the other two regions. The ability to sequence from herbarium specimen material for a given region is ideal for DNA-based identification projects because it vastly increases the size of the sequence library that can be used as a reference when new samples are collected. *trnH-psbA* has also been shown to be easily sequenced across nearly all fern

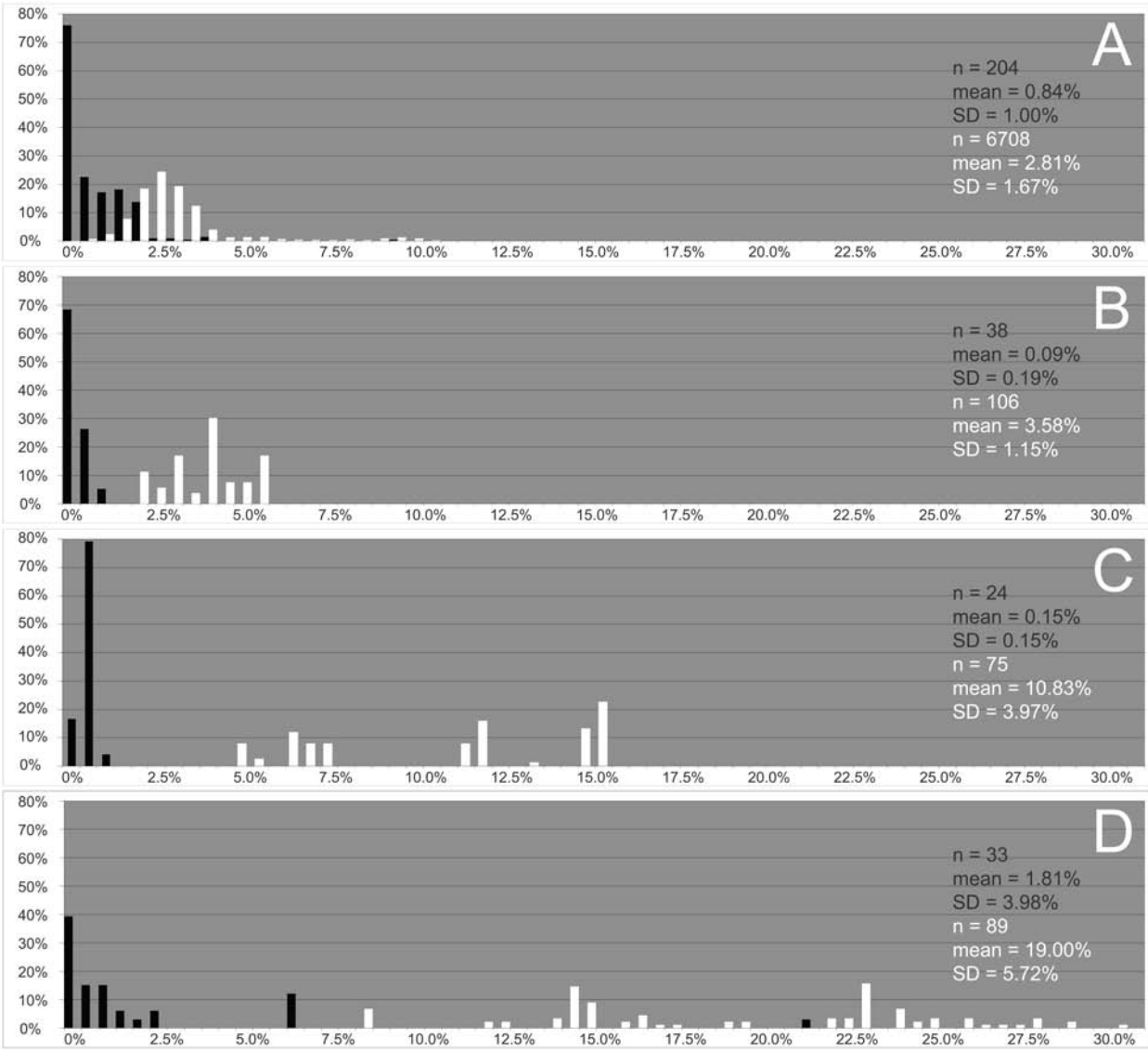

Fig. 4. Intraspecific (black) and interspecific (white) variation for each chloloplast region by sample group. X-axis is variation (uncorrected *p*) with bars corresponding to 0.5% intervals; Y-axis is relative frequency within dataset. A, *rbcL* dataset of broad sampling (GenBank, Moorean, and "additional locality" sequences); B, *rbcL* dataset of restricted sampling (Moorean and "additional locality" sequences only); C, *trnsGG* dataset of restricted sampling; D, *trnH-psbA* dataset of restricted sampling.

families (Small & al., 2005) and a variety of flowering plants (Kress & al., 2005).

A case in point: *Polyphlebium borbonicum*. — During the initial survey of Moorean filmy ferns (Nitta, 2006), one species was found that was previously unrecorded in the literature and could not be easily identified based on morphological characters alone. Examination of herbarium specimens at UC revealed that this species matched in morphology with some, but not all, specimens identified as *Polyphlebium endlicherianum* (C. Presl) Ebihara & K. Iwats. These herbarium specimens sorted into two morphotypes: those with a specialized row of marginal cells along the lamina, and those without (the Moorean specimens matched the later). Since specialized laminar cells, or "false veins," are often an important diagnostic character in filmy fern systematics (Copeland, 1938; Morton, 1968; Dubuisson, 1997; Ebihara & al., 2006), it seemed unlikely that this morphological difference could be attributed to intraspecific variation alone.

This taxonomic puzzle was solved by the addition of DNA sequence data. *rbcL* sequence data of the Moorean specimens and each morphotype of the herbarium specimens were queried against all GenBank *rbcL* data; the Moorean specimens (*JN073* and *JN074*; identical sequences) and the herbarium specimen lacking false veins (UC1797872) matched a *Polyphlebium borbonicum* (Bosch) Ebihara & Dubuisson sequence from Reunion Island (AY175782) at 99.59% and 99.67% respectively, well within the mean intraspecific range of for this gene as determined by the broad sampling barcode analysis (Fig. 4A). The herbarium specimen with false veins (UC1797871) matched a *P. endlicherianum* sequence from New Zealand (AY175787) at 99.01%. *trnH-psbA* confirmed these identifications by revealing average 1.22% and 1.97% variation within putative *P. borbonicum* and *P. endlicherianum* specimens respectively, and average 16.13% variation between the two species; these values fall within the typical intraspecific and interspecific variation observed for this region in the restricted sampling (Fig. 4). (*trnSGG* sequence data could not be obtained from the herbarium specimens, perhaps due to the partially degraded state of their DNA.)

Broader implications. — Results of the present study have several significant implications for DNA-based species identification. First, this study demonstrates that molecular evolutionary rates can vary across lineages even at low (i.e., intrafamilial) taxonomic levels. A significant deceleration in evolutionary rate has been demonstrated in the hymenophylloid clade relative to the trichomanoid clade based on *rbcL* data (Schuettpelz & Pryer, 2006), and this same decrease in rate is also evident in *trnSGG* and *trnH-psbA* (Fig. 5). It is clearly dangerous to make assumptions about evolutionary rates across lineages for a specific DNA region: once consensus on a barcoding region for plants is reached, rates of variation must be confirmed in each lineage of interest before identification can be attempted. Another critical point is that DNA-based identification is only as reliable as the existing taxonomy. High rates of intraspecific variation were observed for known species complexes such as *Hymenophyllum polyanthos* (Sw.) Sw. and *Crepidomanes minutum* (Blume) K. Iwats. (Table 2; Palmer, 2003; Ebihara & al., 2006; Hennequin & al., 2006). Furthermore, both of these species and seven more were found to be paraphyletic or polyphyletic in the Bayesian analysis of the broad sampling dataset (out of 43 species with multiple specimens per species, tree not shown). Fully confident identifications for these species will only be possible after taxonomic revisions have been made based on multiple data of type specimens. Finally, the effects of geographic range of sampling on observed rates of intraspecific variation must be noted. Although rates of intraspecific variation were close to zero for Moorean sequences only (see Results), these rates increased significantly when sequences from additional localities were also included (Table 2). This indicates that sampling across a broad geographic range

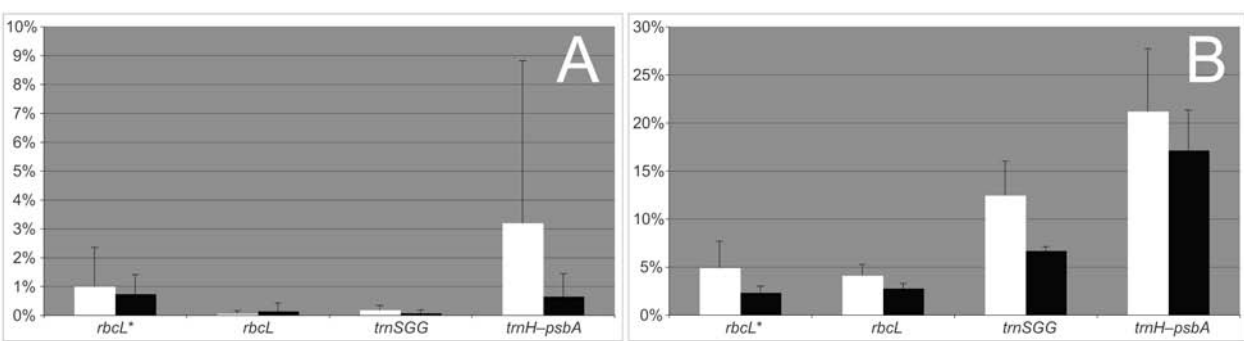

Fig. 5. Variation (uncorrected *p*) for each chloroplast region by clade retrieved by phylogenetic analysis (white = trichomanoid clade, black = hymenophylloid clade). * indicates *rbcL* dataset of broad sampling (GenBank, Moorean, and "additional locality" sequences); other datasets are from restricted sampling (Moorean and "additional locality" sequences only). A, intraspecific variation; B, interspecific variation.

Table 2. Intraspecific variation in each chloroplast region examined by species.

	A. caudatum	***A. dentatum***	***Ca. apiifolia***	***Cr. bipunctatum***
*rbcL**	0.44% ± 0.31% n = 3	0.83% ± 0.65% n = 3	1.13% ± 0.59% n = 6	0.13% ± 0.17% n = 10
rbcL	0.08% n = 1	0.16% n = 1	0.00% n = 1	0.00% n = 6
trnSGG	–	–	–	0.24% ± 0.10% n = 6
trnH-psbA	1.19% n = 1	0.24% n = 1	0.83% n = 1	0.00% n = 3
	Cr. humile	***Cr. kurzii***	***Cr. minutum***	***D. tahitense***
*rbcL**	0.03% ± 0.04% n = 10	0.61% ± 0.53% n = 3	2.19% ± 1.62% n = 10	0.06% ± 0.05% n = 3
rbcL	0.00% n = 6	0.00% n = 1	0.16% ± 0.14% n = 3	0.08% n = 1
trnSGG	0.08% ± 0.03% n = 3	0.23% n = 1	0.04% ± 0.03% n = 3	0.11% n = 1
trnH-psbA	4.14% ± 3.59% n = 3	0.00% n = 1	21.03% n = 1	0.00% n = 1
	P. borbonicum	***H. digitatum***	***H. pallidum***	***H. polyanthos***
*rbcL**	0.28% ± 0.15% n = 6	0.80% ± 0.61% n = 55	0.43% ± 0.36% n = 6	1.09% ± 0.70% n = 36
rbcL	0.16% ± 0.14% n = 3	0.00% n = 6	0.54% ± 0.47% n = 3	0.08% ± 0.09% n = 6
trnSGG	0.34% ± 0.24% n = 3	0.00% n = 3	0.29% n = 1	0.08% ± 0.03% n = 3
trnH-psbA	4.06% ± 3.52% n = 3	0.35% ± 0.39% n = 6	1.47% ± 0.81% n = 6	0.14% ± 0.17% n = 6

* indicates broad sampling (GenBank, Moorean, and "additional locality" *rbcL*), other chloroplast regions are from restricted sampling (Moorean and "additional locality" sequences only). – indicates that an intraspecific pair was not available due to missing data.

is necessary to obtain an accurate measurement of genetic variation within each species.

Examples such as the identification of *P. borbonicum* highlight the synergy that exists between morphological and molecular data and the taxonomic power generated by combining multiple data types. In cases where other data are not available, DNA-based species identification can succeed; however, results of this study show that this is only possible if careful attention is given to variation in evolutionary rates between DNA regions and across lineages. Present results indicate that priority should be placed on the *trnH-psbA* region for the purposes of DNA-based identification in ferns, but sequence data from a variety of gene regions will provide a more resolved picture and will be useful for phylogenetic studies on many levels.

ACKNOWLEDGMENTS

Funding provided by the Gordon and Betty Moore Foundation as part of the Moorea Biocode project. Thanks to Délègation Régionale à la Recherche et à la Technologie, Haut-Commissariat de Republique en Polynésie française for providing collection permits. Thanks to Brent D. Mishler, Andrew Murdock, and Alan Smith for their guidance and comments on drafts; George Roderick and Christopher Meyer provided additional comments on drafts. Thanks to Atsushi Ebihara and Jean-Yves Dubuisson for providing assistance with *Polyphlebium borbonicum* identification and material. The author is grateful to Zachary Hanna, Margaret Groff, Erica Spotswood, Allison Purcell, Elizabeth Perotti, and Victor Wong for assistance in the field and Sonia Nosratinia for assistance in the lab. The present study was conducted as a Senior Honor's Thesis for the BA degree in Integrative Biology at the University of California Berkeley under the supervision of Brent D. Mishler. This paper is contribution 171 from University of California Berkeley's Richard B. Gump South Pacific Research Station, Moorea, French Polynesia.

LITERATURE CITED

Amato, G., Egan, M.G., Schaller, G.B., Baker, R.H., Rosenbaum, H.C., Robichaud, W.G. & DeSalle, R. 1999. The rediscovery of Roosevelt's barking deer (*Muntiacus rooseveltorum*). *J. Mammal.* 80: 639–643.

Baker, C.S. & Palumbi, S.R. 1994. Which whales are hunted—a molecular genetic approach to monitoring whaling. *Science* 265: 1538–1539.

Barrett, R.D.H. & Hebert, P.D.N. 2005. Identifying spiders through DNA barcodes. *Canad. J. Zool.* 83: 481–491.

Blaxter, M.L. 2004. The promise of a DNA taxonomy. *Philos. Trans., Ser. B* 359: 669–679.

Brower, A.V.Z. 2006. Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptes fulgerator* reassessed (Lepidoptera: Hesperiidae). *Syst. Biodiversity* 4: 127–132.

Chase, M.W., Cowan, R.S., Hollingsworth, P.M., Van den Berg, C., Madrinan, S., Petersen, G., Seberg, O., Jorgsensen, T., Cameron, K.M., Carine, M., Pedersen, N., Hedderson, T., Conrad, F., Salazar, G., Richardson, J.E., Hollingsworth, M.L., Barraclough, T.G., Kelly, L. & Wilkinson, M. 2007. A proposal for a standardised protocol to barcode all land plants. *Taxon* 56: 295–299.

Check, E. 2006. Treasure island: pinning down a model ecosystem. *Nature* 439: 378–379.

Copeland, E.B. 1938. Genera Hymenophyllacearum. *Philipp. J. Sci*. 51: 2–110.

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

Dassler, C.L. & Farrar, D.R. 2001. Significance of gametophyte form in long-distance colonization by tropical, epiphytic ferns. *Brittonia* 53: 352–369.

Dayrat, B. 2005. Towards integrative taxonomy. *Biol. J. Linn. Soc*. 85: 407–415.

DeSalle, R. & Birstein, V.J. 1996. PCR identification of black caviar. *Nature* 381: 197–198.

Dubuisson, J.Y. 1997. Systematic relationships within the genus *Trichomanes* sensu lato (Hymenophyllaceae, Filicopsida): cladistic analysis based on anatomical and morphological data. *Bot. J. Linn. Soc*. 123: 265–296.

Dubuisson, J.Y., Hennequin, S., Douzery, E.J.P., Cranfill, R.B., Smith, A.R. & Pryer, K.M. 2003. *rbcL* phylogeny of the fern genus *Trichomanes* (Hymenophyllaceae), with special reference to neotropical taxa. *Int. J. Pl. Sci*. 164: 753–761.

Ebihara, A., Dubuisson, J.Y., Iwatsuki, K., Hennequin, S. & Motomi, I. 2006. A taxonomic revision of Hymenophyllaceae. *Blumea* 51: 1–60.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32: 1792–1797.

Felsenstein, J. 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biol. J. Linn. Soc*. 16: 183–196.

Fosberg, F.R. 1948. Derivation of the flora of the Hawaiian Islands. Pp. 107–119 in: Zimmerman, E.C. (ed.), *Insects of Hawaii*. University of Hawaii Press, Honolulu.

Fox, G.E., Magrum, L.J., Balch, W.E., Wolf, R.S. & Woese, C.R. 1977. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc. Natl. Acad. Sci. U.S.A.* 74: 4537–4541.

Gulamhussein, S. 2000. Distribution and desiccation tolerance of the filmy fern genus *Trichomanes* in two stream valleys on Moorea, French Polynesia. Pp. 76–85 in: Lipps, J. (ed.), *Biology and Geomorphology of Tropical Islands*. Department of Integrative Biology, University of California Berkeley, Berkeley.

Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W. & Hebert, P.D.N. 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. U.S.A.* 103: 968–971.

Hebert, P.D.N., Cywinska, A., Ball, S.L. & DeWaard, J.R. 2003a. Biological identifications through DNA barcodes. *Proc. Roy. Soc. London, Ser. B, Biol. Sci*. 270: 313–321.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. 2003b. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc. Natl. Acad. Sci. U.S.A.* 101: 114812–114817.

Hebert, P.D.N., Ratnasingham, S. & DeWaard, J.R. 2003c. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. Roy. Soc. London, Ser. B, Biol. Sci*. 270: S96–S99.

Hennequin, S., Ebihara, A., Ito, M., Iwatsuki, K. & Dubuisson, J.Y. 2006. New insights into the phylogeny of the genus *Hymenophyllum* s.l. (Hymenophyllaceae): revealing the polyphyly of *Mecodium*. *Syst. Bot*. 31: 271–284.

Huelsenbeck, J.P. & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754–755.

Kress, W.J. & Erickson, D.L. 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(6): e508. doi:10.1371/journal.pone.0000508.

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

Libscomb, D., Platnick, N. & Wheeler, Q.D. 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends Ecol. Evol*. 18: 65–66.

Maddison, D.R. & Maddison, W.P. 2000. *MacClade 4: Analysis of Phylogeny and Character Evolution*, version 4.0. Sinauer, Sunderland.

Meier, R., Shiyang, K., Vaidya, G. & Ng, P.K.L. 2006. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Syst. Biol*. 55: 715–728.

Mengual, X., Stahls, G., Vujic, A. & Marcos-Garcia, A.M. 2006. Integrative taxonomy of Iberian *Merodon* species (Diptera, Syrphidae). *Zootaxa* 1377: 1–26.

Meyer, C.P. & Paulay, G. 2005. DNA barcoding: error rates based on comprehensive sampling. *PLoS Biol*. 3(12): e422. doi:10.1371/journal.pbio.0030422.

Morton, C.V. 1968. The genera, subgenera, and sections of the Hymenophyllaceae. *Contr. U.S. Natl. Herb*. 38: 153–214.

Murdock, A.G. & Smith, A.R. 2003. Pteridophytes of Moorea, French Polynesia, with a New Species, *Tmesipteris gracilis* (Psilotaceae). *Pacific Sci*. 57: 253–265.

Newmaster, S.G., Fazekas, A.J. & Ragupathy, S. 2006. DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach. *Canad. J. Bot*. 84: 335–341.

Nitta, J. 2006. Distribution, ecology, and systematics of the filmy ferns (Hymenophyllaceae) of Moorea, French Polynesia. Pp. 177–199 in: Lipps, J. (ed.), *Biology and Geomorphology of Tropical Islands*. Department of Integrative Biology, University of California Berkeley, Berkeley.

Palmer, D.D. 2003. *Hawai'i's Ferns and Fern Allies*. University of Hawaii Press, Honolulu.

Posada, D. & Crandall, K.A. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.

Prendini, L. 2005. Comment on "identifying spiders through DNA barcodes". *Canad. J. Zool*. 83: 481–491.

Pryer, K.M., Schuettpelz, E., Wolf, P.G., Schneider, H., Smith, A.R. & Cranfill, R.B. 2004. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. *Amer. J. Bot*. 91: 1582–1598.

Pryer, K.M., Smith, A.R., Hunt, J.S. & Dubuisson, J.Y. 2001. *rbcL* data reveal two monophyletic groups of filmy ferns (Filicopsida: Hymenophyllaceae). *Amer. J. Bot*. 88: 1118–1130.

Rambaut, A. & Drummon, A. 2003. Tracer: a program for analysing results from Bayesian MCMC programs such as BEAST & MrBayes. Oxford, U.K. http://evolve.zoo.ox.ac.uk/software.html?id=tracer.

Ranker, T.A., Trapp, P.G., Smith, A.R., Moran, R.C. &

Parris, B.S. 2005. New records of lycophytes and ferns from Moorea, French Polynesia. *Amer. Fern J.* 95: 126–127.

Schneider, H. & Schuettpelz, E. 2006. Identifying fern gametophytes using DNA sequences. *Molec. Ecol. Notes* 6: 989–991.

Schuettpelz, E. & Pryer, K.M. 2006. Reconciling extreme branch length differences: decoupling time and rate through the evolutionary history of filmy ferns. *Syst. Biol.* 55: 485–502.

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E. & Small, R.L. 2005. The totoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Amer. J. Bot.* 92: 142–166.

Simmons, M.P. & Helga, O. 2000. Gaps as characters in sequence-based phylogenetic analysis. *Syst. Biol.* 49: 369–381.

Small, R.L., Lickey, E.B., Shaw, J. & Hauk, W.D. 2005. Amplification of noncoding chloroplast DNA for phylogenetic studies in lycophytes and monilophytes with a comparative example of relative phylogenetic utility from Ophioglossaceae. *Molec. Phylog. Evol.* 36: 509–522.

Smith, A.R. 1972. Comparison of fern and flowering plant distributions with some evolutionary interpretations for ferns. *Biotropica* 4: 4–9.

Smith, A.R., Pryer, K.M., Schuettpelz, E., Korall, P., Schneider, H. & Wolf, P.G. 2006. A classification for extant ferns. *Taxon* 55: 705–731.

Smith, V. 2005. DNA barcoding: perspectives from a "Partnerships for Enhancing Expertise in Taxonomy" (PEET) debate. *Syst. Biol.* 54: 841–844.

Swofford, D.L. 2002. *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*, version 4.0b10. Sinauer, Sunderland.

Taberlet, P., Eric, C., Francios, P., Ludovic, G., Christian, M., Alice, V., Thierry, V., Gerard, C., Christian, B. & Eske, W. 2007. Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* 35: e14. doi: 10.1093/nar/gkl938.

Tate, J. & Simpson, B. 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse orgins of polyploid species. *Syst. Bot.* 28: 723–737.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R. & Hebert, P.D.N. 2005. DNA barcoding Australia's fish species. *Philos. Trans., Ser. B* 360: 1847–1857.

Will, K., Mishler, B.D. & Wheeler, Q.D. 2005. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* 54: 844–851.

Will, K. & Rubinoff, D. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20: 47–55.

Young, N.D. & Healy, J. 2003. GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinformatics* 4: 6.

Appendix 1. Voucher information and GenBank accession numbers for newly sequenced specimens included in this study: species name, locality, collection or herbarium number, *rbcL*, *trnSGG* 5′ end, *trnSGG* 3′ end, *trnH-psbA*. – indicates missing data.

Abrodictyum caudatum (Brack.) Ebihara & K. Iwats., Moorea: Maatea Creek, *JN031*, EU122957, –, –, EU122991; *A. caudatum*, Moorea: Maatea Creek, *JN039*, EU122958, EU123024, EU123050, EU122992; *Abrodictyum dentatum* Cav., Moorea: Mt. Mouaputa, *JN044*, EU122959, –, –, EU122993; *A. dentatum*, Moorea: Maatea Creek, *JN056*, EU122960, EU123025, EU123051, EU122994; *Callistopteris apiifolia* (C. Presl) Copel., Moorea: Maatea Creek, *JN036*, EU122961, –, –, EU122995; *C. apiifolia*, Moorea: Mouaputa, *JN068*, EU122962, EU123026, EU123052, EU122996; *Crepidomanes bipunctatum* (Poir.) Copel., Moorea: Maatea Creek, *JN061*, EU122963, EU123027, EU123053, EU122997; *C. bipunctatum*, Borabora: Mt. Pahia, *JN080*, EU122964, EU123028, EU123054, EU122998; *C. bipunctatum*, Moorea: Cross-island trail, *JN013*, EU338463, EU338471, EU338475, –; *C. bipunctatum*, Moorea: Cross-island trail, *JN014*, EU338464, EU338472, EU338476, EU338483; *Crepidomanes humile* (G. Forst.) Bosch, Moorea: Three Coconuts, *JN001*, EU122965, EU123029, EU123055, –; *C. humile*, Borabora: Mt. Pahia, *JN079*, EU122968, EU123031, EU123057, EU123001; *C. humile*, Moorea: Maatea Creek, *JN022*, EU122966, –, –, EU122999; *C. humile*, Moorea: Maatea Creek, *JN060*, EU122967, EU123030, EU123056, EU123000; *Crepidomanes minutum* (Blume) K. Iwats., Moorea: Mt. Rotui, *JN028*, EU122973, EU123034, EU123060, EU123004; *C. minutum*, Borabora: Mt. Pahia, *JN076*, EU122974, EU123035, EU123061, –; *C. minutum*, Korea, UC1622589, –, –, –, EU338484; *C. minutum*, Moorea: Mt. Mouaputa, *JN041*, EU338465, EU338470, EU338474, –; *Crepidomanes kurzii* (Bedd.) Tagawa & K. Iwats., Moorea: Maatea Creek, *JN021*, EU122969, EU123032, EU123058, EU123002; *C. kurzii*, Moorea: Mt. Mouaputa, *JN052*, EU122970, EU123033, EU123059, EU123003; *Didymoglossum tahitense* (Nadeaud) Ebihara & K. Iwats., Moorea: Three Coconuts, *JN005*, EU122975, EU123036, EU123062, EU123006; *D. tahitense*, Moorea: Marimari, *JN066*, EU122976, EU123037, EU123063, EU123007; *Hymenophyllum digitatum* (Sw.) Fosberg, Moorea: Mt. Rotui, *JN025*, EU122977, –, –, EU338482; *H. digitatum*, Borabora: Mt. Pahia, *JN077*, EU122980, EU123040, EU123066, EU123009; *H. digitatum*, Moorea: Maatea Creek, *JN037*, EU122979, EU123039, EU123065, EU123008; *H. digitatum*, Moorea: Mt. Rotui, *JN027*, EU122978, EU123038, EU123064, EU338481; *Hymenophyllum pallidum* (Blume) Ebihara & K. Iwats., Moorea: Mt. Mouaputa, *JN071*, EU338466, EU123042, EU123068, EU123011; *H. pallidum*, Borabora: Mt. Pahia, *JN075*, EU122982, EU123043, EU123069, EU123012; *H. pallidum*, Caroline Islands, UC1615825, –, –, –, EU123013; *H. pallidum*, Moorea: Mt. Mouaputa, *JN072*, EU338467, –, –, EU338479; *Hymenophyllum polyanthos* (Sw.) Sw., Moorea: Mt. Rotui, *JN006*, EU122983, EU123045, EU123071, EU338480; *H. polyanthos*, Borabora: Mt. Pahia, *JN078*, EU122986, EU123047, EU123073, EU123018; *H. polyanthos*, Moorea: Maatea Creek, *JN034*, EU122984, EU123046, EU123072, EU123016; *H. polyanthos*, Tahiti, *JN062*, EU122985, –, –, EU123017; *Polyphlebium borbonicum* (Bosch) Ebihara & Dubuisson, Moorea: Mt. Mouaputa, *JN073*, EU122988, EU123048, EU123074, EU123019; *P. borbonicum*, Marquesas, *FJ867*, EU338468, EU338469, EU338473, EU338477; *P. borbonicum*, Marquesas, UC1797872, EU348751, –, –, EU123021; *P. borbonicum*, Moorea: Mt. Mouaputa, *JN074*, EU122989, EU123049, EU123075, EU123020; *Polyphlebium endlicherianum* (C. Presl) Ebihara & K. Iwats., Marquesas, UC1797871, EU122990, –, –, EU123023

Appendix 2. GenBank accession numbers for *rbcL* data included in broad sampling. Species name: GenBank accession number.

Abrodictyum asaegrayi (Bosch) Ebihara & K. Iwats: AB257476; *Abrodictyum boninense* Tagawa & K. Iwats.: AB257472; *Abrodictyum brassii* C. Chr.: AB257483; *A. caudatum*: AY175805; *A. dentatum*: AB257477; *Abrodictyum elongatum* (A. Cunn.) Ebihara & K. Iwats.: AY175802; *Abrodictyum flavofuscum* (Bosch) Ebihara & K. Iwats.: AY175804; *Abrodictyum laetum* (Bosch) Ebihara & K. Iwats.: AB257478; *Abrodictyum obscurum* (Blume) Ebihara & K. Iwats.: AB257480; *Abrodictyum pluma* (Hook.) Ebihara & K. Iwats.: AY175803, AB257479; *Abrodictyum rigidum* (Sw.) Ebihara & Dubuisson: AY095108; *Abrodictyum schlechteri* (Brause) Ebihara & K. Iwats.: AB257481; *Abrodictyum strictum* (Menzies ex Hook. & Grev.) Ebihara & K. Iwats.: AB257482; *C. apiifolia*: AB257486, AY175801; *Cephalomanes atrovirens* C. Presl.: AB257484; *Cephalomanes boryanum* Kunze: AB257485; *Cephalomanes javanicum* (Blume) C. Presl.: Y09195; *Crepidomanes aphlebioides* (H. Christ) I.M. Turner: AB257454; *C. bipunctatum*: Y09190; *Crepidomanes christii* (Copel.) Copel.: AB257466; *Crepidomanes fallax* (H. Christ) Ebihara & Dubuisson: AB257459; *Crepidomanes grande* (Copel.) Ebihara & K. Iwats.: AB257487; *C. humile*: AB257470; *Crepidomanes intermedium* (Bosch) Ebihara & K. Iwats.: AY175785; *C. kurzii*: AB257467; *Crepidomanes latealatum* (Bosch) Copel.: AB064297; *C. latealatum*: AB083291; *Crepidomanes latemarginale* (D.C. Eaton) Copel.: AB257468; *Crepidomanes mannii* (Hook.) J.P.Roux: AB257474; *C. minutum*: AB257475, U05625; *Crepidomanes schmidianum* Zenker ex Taschner: AB257465; *Crepidomanes thysanostomum* (Makino) Ebihara & K. Iwats.: AB083294, U05608; *Crepidomanes vitiense* (Baker) Bostock: AB162689; *Crepidomanes walleri* Watts: AB257469; *Didymoglossum bimarginatum* (Bosch) Ebihara & K. Iwats.: AB257494; *Didymoglossum cuspidatum* (Willd.) Ebihara & Dubuisson: AF537122; *Didymoglossum ekmanii* (Wess. Boer) Ebihara & Dubuisson: Y09192; *Didymoglossum erosum* Willd.: AB257495; *Didymoglossum exiguum* (Bedd.) Copel.: AB257488; *Didymoglossum gourlianum* (Grev. ex J. Sm.) Pic. Serm.: Y09194; *Didymoglossum hildebrandtii* (Kuhn) Ebihara & Dubuisson: AY175788; *Didymoglossum hymenoides* (Hedw.) Copel.): AB257489; *Didymoglossum kapplerianum* (Sturm) Ebihara & Dubuisson: AB257496; *Didymoglossum krausii* (Hook. & Grev.) C. Pres: Y09196; *Didymoglossum liberiense* (Copel.) Copel.: AB257490; *Didymoglossum membranaceum* (L.) Vareschi: Y09197; *Didymoglossum motleyi* (Bosch) Ebihara & K. Iwats.: AB257497; *Didymoglossum ovale* E. Fourn.: AB257491; *Didymoglossum pinnatinervium* (Jenman) Pic.Serm.: Y09199; *Didymoglossum punctatum* (Poir.) Desv.): AB257492; *Didymoglossum reptans* (Sw.) C. Presl: AB257493; *D. tahitense*: AB257498; *Hymenophyllum cruentum* Cav.: AB191455, AY095107; *Hymenophyllum acanthoides* (Bosch) Rosenst.: AB064291, AB083282; *Hymenophyllum acutum* (C. Presl) Ebihara & K. Iwats.: AB257473; *Hymenophyllum apiculatum* Mett. ex Kuhn: AF275642; *Hymenophyllum armstrongii* (Baker) Kirk: AB162691, AY095109; *Hymenophyllum australe* Willd.: AB191439; *Hymenophyllum badium* Hook. & Grev.: AB191440; *Hymenophyllum baileyanum* Domin: AB191441, AF275643; *Hymenophyllum barbatum* (Bosch) Baker: AB064287, AB083283; *Hymenophyllum braithwaitei* Ebihara & K. Iwats: AB162686, AB162687; *Hymenophyllum caespitosum* Gaudich.: AB191456, AF275649; *Hymenophyllum caudiculatum* Mart.: AB191442; *Hymenophyllum corrugatum* H. Christ: AB191443; *Hymenophyllum cuneatum* Kunze: AY775401; *Hymenophyllum demissum* G. Forst.: AY775402; *Hymenophyllum deplanchei* Mett. ex Kuhn: AB083284, AB064288; *H. digitatum*: AB162676, AB162677, AB162678, AB162679, AB162680, AB162681, AY095114; *Hymenophyllum dilatatum* G. Forst.: AB191444, AY095111; *Hymenophyllum dimidiatum* Mett. ex Kuhn: AB064289; *Hymenophyllum ferrugineum* Colla: AB191445, AF275644; *Hymenophyllum flabellatum* Labill.: AB083279, AY775403; *Hymenophyllum flexuosum* A. Cunn.: AB217850; *Hymenophyllum frankliniae* Colenso: AB162690; *Hymenophyllum fuciforme* Sw.: AB191446; *Hymenophyllum fucoides* (Sw.) Sw.: U20933; *Hymenophyllum fuscum* (Blume) Bosch: AB064292, AB083285; *Hymenophyllum heimii* Tardieu.: AY775404; *Hymenophyllum hirsutum* (L.) Sw.: AF275645, AY775407; *Hymenophyllum hygrometricum* Poir.: AY095113; *Hymenophyllum inaequale* (Poir.) Desv.: AY095112; *Hymenophyllum javanicum* A. Spreng.: AB191447; *Hymenophyllum lanceolatum* Hook. & Arn.: AF275646; *Hymenophyllum leratii* Rosenst.: AB191448; *Hymenophyllum lyallii* Hook.f.: AB162684, AB162685; *Hymenophyllum marginatum* Hook. & Grev.: AB162692; *Hymenophyllum microcarpum* Desv.: AB083289; *Hymenophyllum mnioides* Baker: AB217849; *Hymenophyllum nitidulum* (Bosch) Ebihara & K. Iwats.: AB162683; *Hymenophyllum oligosorum* Makino: AB064293, AB083280; *Hymenophyllum ooides* F. Muell. & Baker.: AB191449; *H. pallidum*: AB191457; *Hymenophyllum palmatifidum* (Müll.Berol.) Ebihara & K. Iwats.: AB162682; *Hymenophyllum paniense* Ebihara & K. Iwats.: AB083275; *Hymenophyllum pectinatum* Cav.: AB191450, AY095115; *Hymenophyllum pilosissimum* C. Chr.: AB064296, AB083287; *H. polyanthos*: AB064295, AB083276, AF275647, AY775405, AY775406; *Hymenophyllum pulcherrimum* Colenso.: AB191451; *Hymenophyllum rarum* R. Br.: AB217845; *Hymenophyllum nephrophyllum* Ebihara & K. Iwats.: AB083290, U30833; *Hymenophyllum rolandi-principis* Rosenst.: AB064286, AB083286, AY095110; *Hymenophyllum sanguinolentum* (G. Forst.) Sw.: AB191452; *Hymenophyllum scabrum* A. Rich.: AB083278; *Hymenophyllum secundum* Hook. & Grev.: AF275648; *Hymenophyllum sibthorpioides* (Bory ex Willd.) Mett.: AB162688, AY095117; *Hymenophyllum subdimidiatum* Rosenst.: AB064290, AB083281; *Hymenophyllum subobtusum* Rosenst.: AB083288; *Hymenophyllum taeniatum* Copel.: AF275651; *Hymenophyllum tenellum* Jacq.: AB191453, AY095116; *Hymenophyllum tubridgense* (L.) Sm.: Y09203; *Hymenophyllum villosum* Colenso: AB191454; *Hymenophyllum wrightii* Bosch.: AB064294, AB083277; *Polyphlebium angustatum* (Carmich.) Ebihara & Dubuisson: AY175783; *P. borbonicum*: AY175782; *Polyphlebium capillaceum* (L.) Ebihara & Dubuisson: AY175784; *Polyphlebium colensoi* (Hook.) Ebihara & K. Iwats.: AB257456; *Polyphlebium diaphanum* (Kunth) Ebihara & Dubuisson: AB083292, Y09191; *P. endlicherianum*: AY175787; *Polyphlebium exsectum* (Kunze) Ebihara & Dubuisson: AB257458; *Polyphlebium hymenophylloides* (Bosch) Ebihara & Dubuisson: AB257460; *Polyphlebium ingae* (C. Chr.) Ebihara & Dubuisson: AB257461; *Polyphlebium venosum* (R. Br.) Copel.: AY175786; *Polyphlebium vieillardii* (Bosch) Ebihara & K. Iwats.: AB257471; *Trichomanes alatum* Sw.: Y09189; *Trichomanes ankersii* C. Parker ex Hook. & Grev.: AY175800; *Trichomanes arbuscula* Desv.: AY175791; *Trichomanes crinitum* Sw.: AB257501; *Trichomanes crispum* L.: AY175789; *Trichomanes diversifrons* (Bory) Mett.: AY175798; *Trichomanes egleri* P.G. Windisch: AY175797; *Trichomanes elegans* A. Spreng.: AB083295, Y09193; *Trichomanes galeottii* E. Fourn.: AY175794; *Trichomanes holopterum* Kunze: AB257499; *Trichomanes hostmannianum* Klotzsch: AB257500; *Trichomanes lucens* Sw.: AY175792; *Trichomanes mougeotii* Bosch: AY175793; *Trichomanes osmundoides* Poir.: Y09198; *Trichomanes pilosum* Raddi: AY175790; *Trichomanes pinnatum* Hedw.: Y09200;

Appendix 2. Continued.

Trichomanes polypodioides L.: AY175795; *Trichomanes robustum* E. Fourn.: AY175796; *Trichomanes scandens* L.: AB257453, AB257463; *Trichomanes sp.*: Y09202; *Trichomanes trigonum* Desv.: AY175799; *Vandenboschia auriculata* (Blume) Copel.: AB257455; *Vandenboschia birmanicum* Bedd.: AB083293, U05613; *Vandenboschia cyrtotheca* (Hillebr.) Copel.: AB257457; *Vandenboschia davallioides* (Gaudich.) Copel.: U05948; *Vandenboschia johnstonense* (F.M. Bailey) Copel.: AB257462; *Vandenboschia liukiuensis* (Y. Yabe) Copel.: AB196369; *Vandenboschia maxima* (Blume) Copel.: AY175781; *Vandenboschia radicans* (Sw.) Copel.: AB196364, AB196365, AB196366, AB196367, AB196368, AF275650, Y09201; *Vandenboschia rupestris* Ebihara & K. Iwats.: AB257464.