

# DNA Barcode Identification of Podocarpaceae—The Second Largest Conifer Family

Damon P. Little<sup>1\*</sup>, Patrick Knopf<sup>2</sup>, Christian Schulz<sup>2</sup>

**1** Lewis B. and Dorothy Cullman Program for Molecular Systematics, The New York Botanical Garden, Bronx, New York, United States of America, **2** Lehrstuhl für Evolution und Biodiversität der Pflanzen, Ruhr-Universität Bochum, Bochum, Nordrhein-Westfalen, Bundesrepublik Deutschland

## Abstract

We have generated *matK*, *rbcl*, and *nrITS2* DNA barcodes for 320 specimens representing all 18 extant genera of the conifer family Podocarpaceae. The sample includes 145 of the 198 recognized species. Comparative analyses of sequence quality and species discrimination were conducted on the 159 individuals from which all three markers were recovered (representing 15 genera and 97 species). The vast majority of sequences were of high quality ( $B_{30} = 0.596\text{--}0.989$ ). Even the lowest quality sequences exceeded the minimum requirements of the BARCODE data standard. In the few instances that low quality sequences were generated, the responsible mechanism could not be discerned. There were no statistically significant differences in the discriminatory power of markers or marker combinations ( $p = 0.05$ ). The discriminatory power of the barcode markers individually and in combination is low (56.7% of species at maximum). In some instances, species discrimination failed in spite of ostensibly useful variation being present (genotypes were shared among species), but in many cases there was simply an absence of sequence variation. Barcode gaps (maximum intraspecific  $p$ -distance > minimum interspecific  $p$ -distance) were observed in 50.5% of species when all three markers were considered simultaneously. The presence of a barcode gap was not predictive of discrimination success ( $p = 0.02$ ) and there was no statistically significant difference in the frequency of barcode gaps among markers ( $p = 0.05$ ). In addition, there was no correlation between number of individuals sampled per species and the presence of a barcode gap ( $p = 0.27$ ).

**Citation:** Little DP, Knopf P, Schulz C (2013) DNA Barcode Identification of Podocarpaceae—The Second Largest Conifer Family. PLoS ONE 8(11): e81008. doi:10.1371/journal.pone.0081008

**Editor:** Mehrdad Hajibabaei, University of Guelph, Canada

**Received:** April 30, 2013; **Accepted:** October 9, 2013; **Published:** November 27, 2013

**Copyright:** © 2013 Little et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** Funding from the Alfred P. Sloan Foundation (<http://www.sloan.org/>; 2010-6-02) to DPL is gratefully acknowledged. 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: dlittle@nybg.org

## Introduction

Podocarpaceae is a family of evergreen trees and shrubs that are sometimes cultivated as ornamentals in suitably warm climates. In terms of number of species, Podocarpaceae is the second largest family of conifers [1]. Podocarpaceae are often a minor subcanopy component of angiosperm-dominated forests. They are most abundant in the mid- to high-elevation tropics where they thrive on nutrient-poor soils. In addition, Podocarpaceae are found in some unusual low-elevation forest types (e.g. kerangas of Borneo; [2]).

Accurate identification of tropical forest trees, such as Podocarpaceae, is often very difficult. The most easily accessed material is usually sterile. If fertile material is present, it is frequently either inaccessible or detached from the tree making it difficult to convincingly associate the fertile and sterile portions. Although sterile material of Podocarpaceae can usually be identified to genus using phyllotaxis and leaf form [3,4], accurate species identification often requires careful microscopic examination of internal [5–20] and external characteristics [21–27]. Proper use of the existing identification tools requires training in botanical terminology, skill in microtechnique, and familiarity with Podocarpaceae.

Species of Podocarpaceae are of conservation concern primarily as a result of small population sizes and limited available habitat. Twenty-seven Podocarpaceae species are included in the Inter-

national Union for the Conservation of Nature (IUCN; [28]) red list under the categories of vulnerable (10 species), endangered (14 species), and critically endangered (three species). Two species are included in the appendices of the Convention on International Trade in Endangered Species (CITES; [29]): *Podocarpus parlatorei* is listed in Appendix I (trade is not allowed) and *Po. neriiifolius* is listed in Appendix III (trade with, some limitations, is allowed).

Podocarpaceae have a minor role in commerce. *Nageia nagi*, when labeled as Asian bayberry, can legally be sold in the United States of America as an herbal dietary supplement [30]. The seeds are processed into an edible oil that is also used in manufacturing [31]. The young leaves are also edible, but not typically consumed [32]. The conspicuous fleshy reproductive structures (receptacles or epimatium) of *Afrocarpus falcatus*, *Dacrycarpus dacrydioides*, *Dacrydium cupressinum*, *Po. elatus*, *Po. macrophyllum*, *Po. totara*, and *Prumnopitys taxifolia* are eaten either raw or cooked [32].

Although their use is currently very limited, Podocarpaceae are known to have medicinal properties that benefit humans and animals [33,34]. The receptacles and leaves contain a variety of bio-active compounds such as antioxidants, nordi-terpenes, podocarpic acid, and tatarol [33,35,36]. Some of these compounds have antimicrobial, fungistatic, or bacteriostatic properties [33,37,38]. Other compounds have cytotoxic properties that may be useful in destroying cancer [39–43].

The rarity of large uniform stands coupled with a slow rate of growth for most species makes harvest of Podocarpaceae wood generally unsustainable [2]. The growth rate of *Po. totara* may however accommodate sustainable harvest [44]. Relative scarcity results in a meager international trade—primarily originating from New Zealand and South Africa [2,45–47]. Timber from Podocarpaceae, referred to as ‘podo’ (or ‘yellow yew’) in commerce, has straight even grain. The wood of some species is brittle when worked and not particularly durable outdoors [46,47]. Wood from *Po. totara* is durable and highly amenable to industrial machining [44]. Wood from *Lepidothamnus intermedius*, *Manoao colensoi*, and *Pr. taxifolia* is very rot resistant [48]. Timber of *Ma. colensoi* and *Le. intermedius* have long been used for railway ties [48]. In addition, the wood of some Podocarpaceae species is highly insect resistant (e.g. *Af. gracilior* [49], *Po. hallii* [50], *Po. macrophyllus* [51], and *Po. nivalis* [50]).

A reference library of Podocarpaceae DNA barcodes will allow researchers unfamiliar with the family’s morphology and anatomy to make accurate identifications. We hope that DNA barcodes will permit foresters, ecologists, conservationists, customs authorities, etc. to make accurate biodiversity inventories and to monitor trade in threatened and endangered Podocarpaceae species so that future conservation and management decisions can be based on sound data.

We aim to generate and evaluate a DNA barcode reference library for Podocarpaceae. The library will be assessed both by the quality of the constituent sequences and the degree to which observed sequence variation unambiguously distinguishes Podocarpaceae species from one another.

## Materials and Methods

We sampled 320 individuals representing all of the 18 extant genera of Podocarpaceae (including *Phyllocladus*). Our sample included 145 of the 198 recognized species (73.2%; [1]). Between 1 and 9 individuals per species were sequenced (median = 2; IQR = 1–3). All samples were expert-identified using a combination of morphology and leaf anatomy [20]. Voucher information is in Dataset S1.

## Molecular techniques

DNA was extracted from herbarium specimens or silica-dried tissue using the Qiagen DNeasy96 kit. The manufacturer’s protocol was modified for herbarium specimens: instead of the recommended incubation, homogenized tissue was digested with 30  $\mu$ L (20  $\mu$ g/ $\mu$ L stock) of proteinase K in 400  $\mu$ L AP1 (supplied by the manufacturer) and 1  $\mu$ L of DX (supplied by the manufacturer) at 42°C for 24 hours with slow mixing (60 rotations per minute).

The Polymerase Chain Reaction (PCR) was used to amplify *matK* in a 15  $\mu$ L volume containing: 20 mM tris pH 8.8, 10 mM KCl, 10 mM (NH<sub>4</sub>)<sub>2</sub> SO<sub>4</sub>, 2 mM MgSO<sub>4</sub>, 0.1% (v/v) Triton X-100, 5% (w/v) sucrose, 0.025% (w/v) cresol red, 0.025  $\mu$ g/ $\mu$ L BSA, 0.2 mM dNTPs, 1  $\mu$ M of Gym\_F1A (5’-ATY GYR CTT TTA TGT TTA CAR GC-3’; [53]), 1  $\mu$ M of Gym\_R1A (5’-TCA YCC GGA RAT TTT GGT TCG-3’; [53]), 0.5 units *Taq*, and 0.5  $\mu$ L genomic DNA. The reaction mixture was incubated for 150 sec at 95°C, cycled 35 times (30 sec at 95°C, 60 sec at 52°C, 40 sec at 72°C) and then incubated at 72°C for 10 minutes.

PCR amplification of nrITS2 was similar to that of *matK* except primer annealing was carried out for 30 sec at 58°C rather than 60 sec at 52°C and primers S2F (5’-ATG CGA TAC TTG GTG TGA AT-3’; [53]) and S3R (5’-GAC GCT TCT CCA GAC TAC AAT-3’ [53]) replaced Gym\_F1A and Gym\_R1A.

Unused primers and dNTPs were neutralized using ExoSAP-IT (USB). PCR products were bidirectionally sequenced, using the amplification primers, with BigDye v3.1 (Life Technologies) at the High-Throughput Genomics Unit (University of Washington). PCR amplification and sequencing of *rbcL* was described previously [54].

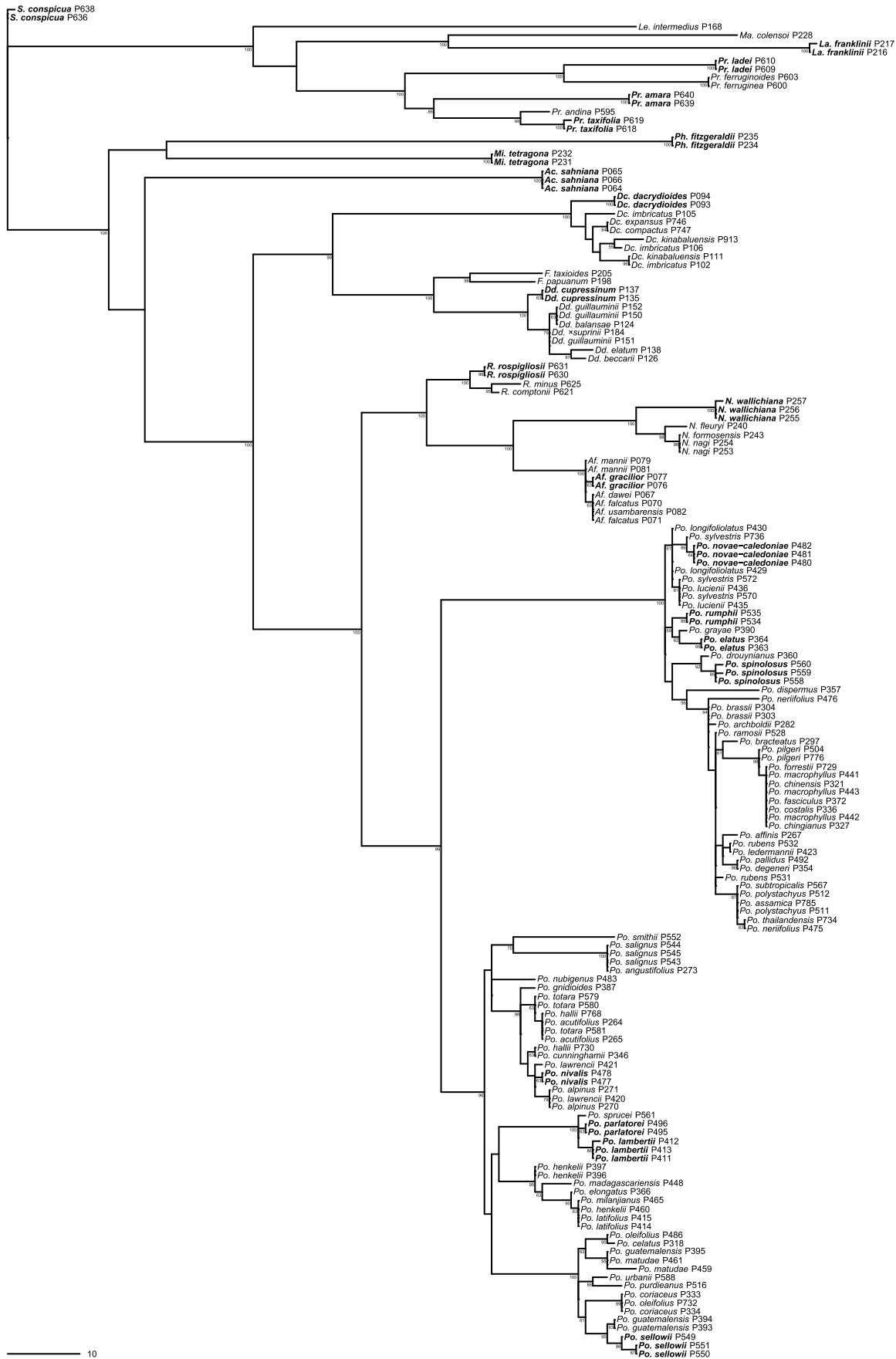
## Data analysis

Bases were called and quality values (QV) assigned using KB 1.4 (Life Technologies). Sequencer 4.1 (Gene Codes) was used to construct sequence contigs, trim contigs to a uniform beginning/end (priming sites were excluded), and resolve differences between sequencing reads. Sequences of *matK* and *rbcL* were checked for stop codons and frameshift mutations. To identify potential contaminants, all sequences were queried against GenBank using BLAST 2.2.26 [55]. Only hits with an e-value of 10<sup>-20.0</sup> or less were retained. Additional contaminants were identified by aligning each marker with MUSCLE 3.8 [56], coding the resulting indels using ‘simple indel coding’ [57,58], and resampling the resulting matrix 1000 times using the jackknife [59] procedure. For each resampled matrix, the search for optimal trees was conducted in TNT 1.1 [60]. The search consisted of ten random addition replicates with five trees held in memory per replicate and SPR followed by TBR (BB) branch swapping. The strict consensus tree from each resampled matrix was used to calculate the jackknife tree.

In order to make meaningful comparisons across markers, we only analyzed data from specimens for which *matK*, *rbcL*, and nrITS2 were able to be sequenced—referred to as ‘complete samples’ hereafter. Sequences from specimens that could not be definitively identified using morphology/anatomy were excluded.

Sequence quality was assessed using the barcode quality index (*B*; [61]) with the acceptable quality threshold (*q*) set to 30 (an average of one error per thousand sequenced bases). The expected coverage (*x*) was set to 2. The contig size (*c*) was set to the observed size. Linguistic complexity (*LC*; [62]), a measure of sequence repetitiveness, was calculated for each sequence, using COMPLEX 6.1.0 [63] with window size set to 100 bases, step size set to 1 base, minimum pattern size set to 3 bases, and maximum pattern size set to 6 bases. The threshold for significant increase in homopolymer (mononucleotide repeats) induced PCR artifacts has been empirically determined to be eight bases [64]—thus sequences with homopolymers, eight bases or longer, were identified. Statistical differences, in sequence quality, linguistic complexity, and homopolymer frequency among markers were evaluated with Scheffé’s test [65–67] at *p*=0.05 using the Gaussian distribution. Correlations between sequence quality and linguistic complexity as well as sequence quality and homopolymer frequency were measured by Spearman’s rank correlation tests [67,68].

TNT 1.1 [60] was used to analyze phylogenetic relationships among complete samples. Each marker was aligned with MUSCLE 3.8 [56], indels were coded using ‘simple indel coding’ [57,59], and markers were combined by concatenation [69]. The resulting matrix was searched for optimal trees using 1000 random addition replicates: for each replicate two trees were held in memory, SPR branch swapping was followed by TBR (BB) branch swapping and a 200 iteration ratchet [70] perturbing 8% of the characters per iteration (4% up weighted, 4% down weighted). Clade support was assessed by 10,000 jackknife resamplings [59]. For each resampled matrix, the search for optimal trees consisted of ten random addition replicates with five trees held in memory per replicate and SPR followed by TBR (BB) branch swapping. The jackknife frequency of each clade in the strict consensus of the



10

**Figure 1. Phylogenetic relationships among complete samples.** Strict consensus of 3600 most parsimonious trees (L=1205; CI=0.59; RI=0.93; all tree statistics exclude uninformative nucleotide positions) obtained from the simultaneous analysis of *matK*, *rbcL*, and nrITS2 sequence data. Numbers at nodes indicated jackknife support above 50%. Species that can be distinguished from all other species using the 'least inclusive clade' method are in boldface (the least inclusive clade method cannot be applied to species with only one sample). Genera have been abbreviated: *Ac.* = *Acmopyle*, *Af.* = *Afrocarpus*, *Dc.* = *Dacrycarpus*, *Dd.* = *Dacrydium*, *F.* = *Falcatifolium*, *La.* = *Lagarostrobos*, *Le.* = *Lepidothamnus*, *Ma.* = *Manoao*, *Mi.* = *Microcachrys*, *N.* = *Nageia*, *Ph.* = *Pherosphaera*, *Po.* = *Podocarpus*, *Pr.* = *Prumnopitys*, *R.* = *Retrophyllum*, and *S.* = *Saxegothaea*. Sample codes correspond to those used in Dataset S1. doi:10.1371/journal.pone.0081008.g001

original matrix was calculated with SUMTREES 3.3.1 [71] using the strict consensus tree from each resampled matrix [67,72]. Trees were rooted following [54]. Tree-based species discrimination was assessed using the 'least inclusive clade' method [73].

Species discrimination was calculated using BRONX 2.0 [74,75]. Discrimination success would be overestimated if the reference database just included sequences in the complete sample—thus a BRONX reference database was constructed from all sequences for each marker and marker combination (Dataset S1). To calculate species discrimination, sequences of each complete sample were queried against the reference database. Species were considered distinct if all queries for a given species returned only sequences belonging to that species. The binomial distribution, with each species considered an independent test, was used to compute 95% confidence intervals [67,76,77]. Differences in species discrimination among markers and marker combinations were quantified using Scheffé's test [65–67] at  $p=0.05$ . The binomial distribution was used for tests of species discrimination and the Gaussian distribution was used to test if the number of species conflated when identification failed varied among markers.

Relative variation within and among species—the 'barcode gap' [78]—was quantified by comparing pairwise distances for complete samples. Each pair of sequences in the complete sample was aligned separately with MUSCLE 3.8 [56] and the number of unambiguous nucleotide differences was divided by the total number of aligned positions to calculate the edit distance (uncorrected p-distance; [79]). To minimize sampling and analytic artifacts, the maximum intraspecific distance was compared to the minimum interspecific distance for each species [80]. For each marker, the frequency of barcode gap (maximum intraspecific > minimum interspecific) occurrence was assessed using the binomial distribution and Scheffé's test [65–67] at  $p=0.05$ . The point-biserial correlation coefficient was used to examine the relationship between number of samples per species and the occurrence of a barcode gap [67,81]. Sequences from all three markers were used simultaneously with McNemar's test [67,82] to measure the correlation between the occurrence of a barcode gap and whether or not a species can be consistently distinguished from all other species using diagnostic nucleotide positions.

## Results and Discussion

In total, 281 *matK*, 202 *rbcL*, and 212 nrITS2 finished sequences were generated (Dataset S1). BLAST [55] queries indicate that the newly generated sequences are consistent with other samples of Podocarpaceae deposited in GenBank (data not shown). Phylogenetic arrangement of genera and species is roughly consistent with previous molecular phylogenetic studies (Figure 1; [54,83–86]). Sequences derived from individuals of the same species are always in close phylogenetic proximity, but in some cases the sequences do not form a monophyletic group (sensu [87]). Sequences of some morphological/anatomical species are unambiguously polyphyletic (sensu [87]; e.g. *Podocarpus oleifolius*, Figure 1). Mismatches between morphological/anatomical species circumscription and barcode sequences warrant further investigation as they may

indicate the presence of cryptic species, introgression, or ancestral polymorphism followed by incomplete lineage sorting. Together the BLAST and phylogenetic contaminate screens indicate that the sequences generated are indeed Podocarpaceae and that no PCR artifacts or errors in sample handling could be detected.

Newly generated sequences of *matK* vary from 760 to 775 bp (median = 769; IQR = 769–769), sequences of *rbcL* are uniformly 607 bp, and sequences of nrITS2 vary from 420 to 435 bp (median = 425; IQR = 425–425). The multiple sequence alignment of *matK* results in six indels—3 bp (two indels), 6 bp (three indels), and 9 bp (one indel), respectively. The 20 nrITS2 indels resulting from multiple sequence alignment range from 1 to 17 bp (median = 1; IQR = 1–2).

Of the 320 individuals sequenced, finished *matK*, *rbcL*, and nrITS2 sequences were generated for 159 individuals. These samples, representing 15 of the 18 extant genera (83.3%; [1]; *Halocarpus*, *Parasitaxus*, and *Phyllocladus* are not included in the complete sample) and 97 of the 198 recognized species (48.9%; [1]), were analyzed for sequence quality, linguistic complexity, species discrimination, and barcode gaps. The complete sample set contained between 1 and 3 individuals per species (median = 1; IQR = 1–2; Table 1). In total, there were 95 distinct *matK* sequence types, 70 *rbcL* sequence types, and 81 nrITS2 sequence types. The complete sample contained 71 (74.7%) *matK* sequence types, 61 (87.1%) *rbcL* sequence types, and 65 (80.2%) nrITS2 sequence types for a combined 90 distinct multilocus genotypes.

### Sequence quality and complexity

Sequence quality, as measured by  $B_{30}$  [61], ranged from 0.775 to 0.989 for *matK* (median = 0.967; IQR = 0.960–0.975), 0.596 to 0.951 for *rbcL* (median = 0.938; IQR = 0.929–0.944), and 0.671 to 0.933 for nrITS2 (median = 0.924; IQR = 0.919–0.927; Figure 2). The vast majority of sequences were of high quality: across all markers, 93.5% of the positions in the median sequence were assigned a quality value of 30 or greater—indicating that few, if any, of the finished sequences contain erroneous base calls. Although differences in sequence quality among markers was statistically significant ( $p=0.05$ ; *matK* > *rbcL* > nrITS2), even the lowest quality sequences exceeded the minimum requirements of the BARCODE data standard (version 2.3; [88])—thus the statistical differences observed are not particularly meaningful in practice.

Published  $B_{30}$  values for sequences generated using different primer sets are not directly comparable to those reported here because the primer sets define (slightly) different marker regions. Although not comparable in all cases, the median Podocarpaceae sequence is of higher quality than the average sequence reported for angiosperms across all three markers [61,89,90]. The largest comparable difference between literature reports and newly generated Podocarpaceae sequences was observed in nrITS2 (0.829 versus 0.924; [89]). The high quality of Podocarpaceae *matK* sequences is notable, but the gymnosperm specific primers [52] used to generate the *matK* sequences make direct comparisons to published values for angiosperms tenuous.

**Table 1.** Diagnostic barcode variation for complete samples of Podocarpaceae.

species	n	discriminatory success				multilocus barcode gap	most frequently conflated species (ordered by frequency)
		matK	rbcl	nrITS2	combined		
<i>Ac. sahniana</i>	3	+	+	+	+	1	—
<i>Af. dawei</i>	1	—	—	—	—	0	<i>Af. gracilior</i> , <i>Af. mannii</i> , <i>Af. falcatus</i> , et al.
<i>Af. falcatus</i>	2	—	—	—	—	0	<i>Af. gracilior</i> , <i>Af. mannii</i> , <i>Af. usambarensis</i> , et al.
<i>Af. gracilior</i>	2	—	—	+	+	0	<i>Af. falcatus</i> , <i>Af. usambarensis</i> , <i>Af. dawei</i> , et al.
<i>Af. mannii</i>	2	—	—	—	+	0	<i>Af. falcatus</i> , <i>Af. usambarensis</i> , <i>Af. dawei</i> , et al.
<i>Af. usambarensis</i>	1	—	—	—	—	0	<i>Af. gracilior</i> , <i>Af. mannii</i> , <i>Af. falcatus</i> , et al.
<i>Dc. compactus</i>	1	—	—	—	—	0	<i>Dc. kinabaluensis</i> , <i>Dc. imbricatus</i> , <i>Dc. expansus</i>
<i>Dc. dacrydioides</i>	2	+	+	+	+	1	—
<i>Dc. expansus</i>	1	—	—	—	—	0	<i>Dc. kinabaluensis</i> , <i>Dc. imbricatus</i> , <i>Dc. compactus</i>
<i>Dc. imbricatus</i>	3	—	—	+	+	0	<i>Dc. expansus</i> , <i>Dc. compactus</i> , <i>Dc. kinabaluensis</i>
<i>Dc. kinabaluensis</i>	2	—	—	—	+	0	<i>Dc. expansus</i> , <i>Dc. compactus</i> , <i>Dc. imbricatus</i>
<i>Dd. balansae</i>	1	—	—	—	—	0	<i>Dd. araucarioides</i> , <i>Dd. cupressinum</i> , <i>Dd. nausoriense</i> , et al.
<i>Dd. beccarii</i>	1	—	+	+	+	1	<i>Dd. gracile</i> , <i>Dd. xanthandrum</i>
<i>Dd. cupressinum</i>	2	+	—	+	+	1	<i>Dd. guillauminii</i> , <i>Dd. araucarioides</i> , <i>Dd. balansae</i> , et al.
<i>Dd. elatum</i>	1	+	+	+	+	1	—
<i>Dd. guillauminii</i>	3	—	—	—	—	0	<i>Dd. araucarioides</i> , <i>Dd. cupressinum</i> , <i>Dd. nausoriense</i> , et al.
<i>Dd. xsuprinii</i>	1	—	—	—	—	0	<i>Dd. araucarioides</i> , <i>Dd. cupressinum</i> , <i>Dd. nausoriense</i> , et al.
<i>F. papuanum</i>	1	+	+	+	+	1	—
<i>F. taxioides</i>	1	+	+	+	+	1	—
<i>La. franklinii</i>	2	+	+	+	+	1	—
<i>Le. intermedius</i>	1	+	+	+	+	1	—
<i>Ma. colensoi</i>	1	+	—	+	+	1	<i>Le. intermedius</i>
<i>Mi. tetragona</i>	2	+	+	+	+	1	—
<i>N. fleuryi</i>	1	+	+	+	+	1	—
<i>N. formosensis</i>	1	—	—	—	—	0	<i>N. motleyi</i> , <i>N. nagi</i>
<i>N. nagi</i>	2	—	—	—	—	0	<i>N. motleyi</i> , <i>N. formosensis</i>
<i>N. wallichiana</i>	3	+	—	+	+	1	<i>N. motleyi</i>
<i>Ph. fitzgeraldii</i>	2	+	+	+	+	1	—
<i>Po. acutifolius</i>	2	—	—	—	—	0	<i>Po. cunninghamii</i> , <i>Po. totara</i> , <i>Po. hallii</i>
<i>Po. affinis</i>	1	—	+	+	+	1	<i>Po. degeneri</i> , <i>Po. rubens</i> , <i>Po. insularis</i> , et al.
<i>Po. alpinus</i>	2	—	—	—	—	0	<i>Po. lawrencei</i> , <i>Po. gnidioides</i> , <i>Po. nivalis</i> , et al.
<i>Po. angustifolius</i>	1	—	—	—	—	0	<i>Po. salignus</i>
<i>Po. archboldii</i>	1	+	—	—	+	1	<i>Po. drouynianus</i> , <i>Po. polystachyus</i> , <i>Po. thailandensis</i> , et al.
<i>Po. assamica</i>	1	—	—	—	—	0	<i>Po. drouynianus</i> , <i>Po. deflexus</i> , <i>Po. insularis</i> , et al.
<i>Po. bracteatus</i>	1	—	—	—	+	1	<i>Po. neriifolius</i> , <i>Po. pseudobracteatus</i>
<i>Po. brassii</i>	2	—	—	—	+	1	<i>Po. drouynianus</i> , <i>Po. thailandensis</i> , <i>Po. deflexus</i> , et al.
<i>Po. celatus</i>	1	—	—	—	+	1	<i>Po. guatemalensis</i> , <i>Po. coriaceus</i> , <i>Po. sellowii</i> , et al.
<i>Po. chinensis</i>	1	—	—	—	—	0	<i>Po. nakaii</i> , <i>Po. pilgeri</i> , <i>Po. annamiensis</i> , et al.
<i>Po. chingianus</i>	1	—	—	—	—	0	<i>Po. nakaii</i> , <i>Po. pilgeri</i> , <i>Po. annamiensis</i> , et al.
<i>Po. coriaceus</i>	2	—	—	—	—	0	<i>Po. guatemalensis</i> , <i>Po. trinitensis</i> , <i>Po. celatus</i> , et al.
<i>Po. costalis</i>	1	—	—	—	—	0	<i>Po. nakaii</i> , <i>Po. pilgeri</i> , <i>Po. annamiensis</i> , et al.
<i>Po. cunninghamii</i>	1	—	—	—	—	0	<i>Po. totara</i> , <i>Po. acutifolius</i> , <i>Po. hallii</i>
<i>Po. degeneri</i>	1	—	—	—	—	0	<i>Po. affinis</i> , <i>Po. insularis</i> , <i>Po. ledermannii</i> , et al.
<i>Po. dispermus</i>	1	+	+	+	+	1	—
<i>Po. drouynianus</i>	1	—	+	—	+	1	<i>Po. polystachyus</i> , <i>Po. thailandensis</i> , <i>Po. insularis</i> , et al.
<i>Po. elatus</i>	2	+	+	+	+	1	—
<i>Po. elongatus</i>	1	+	+	—	+	1	<i>Po. polystachyus</i> , <i>Po. milanjanus</i> , <i>Po. henkelii</i> , et al.
<i>Po. fasciculus</i>	1	—	—	—	—	0	<i>Po. nakaii</i> , <i>Po. pilgeri</i> , <i>Po. annamiensis</i> , et al.

Table 1. Cont.

species	n	discriminatory success				multilocus barcode gap	most frequently conflated species (ordered by frequency)
		matK	rbcl	nrITS2	combined		
<i>Po. forrestii</i>	1	–	–	–	–	0	<i>Po. nakaii</i> , <i>Po. pilgeri</i> , <i>Po. annamiensis</i> , et al.
<i>Po. gnidioides</i>	1	+	–	+	+	1	<i>Po. alpinus</i> , <i>Po. lawrencii</i>
<i>Po. grayae</i>	1	+	–	+	+	1	<i>Po. lucienii</i> , <i>Po. rumphii</i> , <i>Po. sylvestris</i> , et al.
<i>Po. guatemalensis</i>	3	–	–	–	–	0	<i>Po. tepuiensis</i> , <i>Po. brasiliensis</i> , <i>Po. coriaceus</i> , et al.
<i>Po. hallii</i>	2	–	–	–	–	0	<i>Po. totara</i> , <i>Po. acutifolius</i> , <i>Po. cunninghamii</i>
<i>Po. henkelii</i>	3	–	–	–	–	0	<i>Po. polystachyus</i> , <i>Po. elongatus</i> , <i>Po. milanjanus</i> , et al.
<i>Po. lambertii</i>	3	+	+	+	+	1	—
<i>Po. latifolius</i>	2	–	–	–	–	0	<i>Po. polystachyus</i> , <i>Po. elongatus</i> , <i>Po. milanjanus</i> , et al.
<i>Po. lawrencii</i>	2	–	–	–	–	0	<i>Po. lawrencei</i> , <i>Po. gnidioides</i> , <i>Po. nivalis</i> , et al.
<i>Po. ledermannii</i>	1	–	–	–	–	0	<i>Po. polystachyus</i> , <i>Po. thailandensis</i> , <i>Po. decipiens</i> , et al.
<i>Po. longifoliolatus</i>	2	–	–	–	+	1	<i>Po. rumphii</i> , <i>Po. grayae</i> , <i>Po. lucienii</i> , et al.
<i>Po. lucienii</i>	2	–	–	–	–	0	<i>Po. rumphii</i> , <i>Po. grayae</i> , <i>Po. decumbens</i> , et al.
<i>Po. macrophyllus</i>	3	–	–	–	–	0	<i>Po. nakaii</i> , <i>Po. pilgeri</i> , <i>Po. annamiensis</i> , et al.
<i>Po. madagascariensis</i>	1	+	+	+	+	1	—
<i>Po. matudae</i>	2	–	–	–	–	0	<i>Po. oleifolius</i> , <i>Po. guatemalensis</i>
<i>Po. milanjanus</i>	1	–	–	–	–	0	<i>Po. polystachyus</i> , <i>Po. elongatus</i> , <i>Po. henkelii</i> , et al.
<i>Po. neriifolius</i>	2	–	–	–	–	0	<i>Po. drouynianus</i> , <i>Po. insularis</i> , <i>Po. ledermannii</i> , et al.
<i>Po. nivalis</i>	2	–	+	–	+	1	<i>Po. lawrencei</i> , <i>Po. totara</i> , <i>Po. acutifolius</i> , et al.
<i>Po. novae-caledoniae</i>	3	–	–	–	+	1	<i>Po. sylvestris</i> , <i>Po. beecherae</i>
<i>Po. nubigenus</i>	1	–	+	+	+	1	<i>Po. atjehensis</i>
<i>Po. oleifolius</i>	2	–	–	–	–	0	<i>Po. trinitensis</i> , <i>Po. salicifolius</i> , <i>Po. rusbyi</i> , et al.
<i>Po. pallidus</i>	1	–	–	–	–	0	<i>Po. affinis</i> , <i>Po. insularis</i> , <i>Po. ledermannii</i> , et al.
<i>Po. parlatorei</i>	2	–	–	+	+	1	<i>Po. transiens</i> , <i>Po. sprucei</i>
<i>Po. pilgeri</i>	2	–	+	–	+	1	<i>Po. nakaii</i> , <i>Po. costalis</i> , <i>Po. fasciculus</i> , et al.
<i>Po. polystachyus</i>	2	–	–	–	+	0	<i>Po. drouynianus</i> , <i>Po. deflexus</i> , <i>Po. insularis</i> , et al.
<i>Po. purdieanus</i>	1	+	+	+	+	1	—
<i>Po. ramosii</i>	1	–	–	–	+	1	<i>Po. thailandensis</i> , <i>Po. insularis</i> , <i>Po. assamica</i> , et al.
<i>Po. rubens</i>	2	–	–	–	–	0	<i>Po. drouynianus</i> , <i>Po. polystachyus</i> , <i>Po. thailandensis</i> , et al.
<i>Po. rumphii</i>	2	–	–	+	+	1	<i>Po. lucienii</i> , <i>Po. sylvestris</i> , <i>Po. grayae</i> , et al.
<i>Po. salignus</i>	3	–	–	–	–	0	<i>Po. angustifolius</i>
<i>Po. sellowii</i>	3	+	–	–	+	1	<i>Po. tepuiensis</i> , <i>Po. celatus</i> , <i>Po. oleifolius</i> , et al.
<i>Po. smithii</i>	1	+	+	+	+	1	—
<i>Po. spinulosus</i>	3	–	–	+	+	1	<i>Po. dispermus</i> , <i>Po. drouynianus</i> , <i>Po. rostratus</i> , et al.
<i>Po. sprucei</i>	1	–	–	–	+	1	<i>Po. glomeratus</i> , <i>Po. lambertii</i> , <i>Po. transiens</i> , et al.
<i>Po. subtropicalis</i>	1	–	–	–	–	0	<i>Po. drouynianus</i> , <i>Po. deflexus</i> , <i>Po. insularis</i> , et al.
<i>Po. sylvestris</i>	3	–	–	–	–	0	<i>Po. rumphii</i> , <i>Po. grayae</i> , <i>Po. beecherae</i> , et al.
<i>Po. thailandensis</i>	1	–	–	–	–	0	<i>Po. drouynianus</i> , <i>Po. insularis</i> , <i>Po. ledermannii</i> , et al.
<i>Po. totara</i>	3	–	–	–	–	0	<i>Po. cunninghamii</i> , <i>Po. acutifolius</i> , <i>Po. hallii</i>
<i>Po. urbanii</i>	1	+	–	+	+	1	<i>Po. guatemalensis</i> , <i>Po. celatus</i> , <i>Po. oleifolius</i> , et al.
<i>Pr. amara</i>	2	+	+	+	+	1	—
<i>Pr. andina</i>	1	+	–	+	+	1	<i>Pr. exigua</i> , <i>Pr. montana</i>
<i>Pr. ferruginea</i>	1	–	–	–	–	0	<i>Pr. ferruginoides</i>
<i>Pr. ferruginoides</i>	1	–	–	+	+	0	<i>Pr. ferruginea</i>
<i>Pr. ladei</i>	2	+	+	+	+	1	—
<i>Pr. taxifolia</i>	2	+	+	+	+	1	—
<i>R. comptonii</i>	1	+	–	–	+	1	<i>R. minus</i> , <i>R. rospigliosii</i> , <i>R. vitiense</i>
<i>R. minus</i>	1	–	–	+	+	1	<i>R. comptonii</i>
<i>R. rospigliosii</i>	2	+	+	–	+	1	<i>R. comptonii</i> , <i>R. vitiense</i>

Table 1. Cont.

species	n	discriminatory success				multilocus barcode gap	most frequently conflated species (ordered by frequency)
		matK	rbcL	nrITS2	combined		
<i>S. conspicua</i>	2	+	+	+	+	1	—

Genera have been abbreviated: Ac. = Acmopyle, Af. = Afrocarpus, Dc. = Dacrycarpus, Dd. = Dacrydium, F. = Falcatifolium, La. = Lagarostrobos, Le. = Lepidothamnus, Ma. = Manoa, Mi. = Microcachrys, N. = Nageia, Ph. = Pherosphaera, Po. = Podocarpus, Pr. = Prumnopitys, R. = Retrophyllum, and S. = Saxegothaea. Species that can be consistently distinguished from all other species are indicated by a '+' while those that are conflated with other species are indicated with a '-'. The presence of a barcode gap is indicated with '1', absence of a barcode gap with '0'.  
doi:10.1371/journal.pone.0081008.t001

Linguistic complexity is a measure of the number of repeated 'words' in a sequence (words 3–6 bp were examined in this case; [62]). Sequences of *matK*, *rbcL*, and nrITS2 have statistically distinct linguistic complexity ( $p=0.05$ ) with *matK* being the simplest (median = 0.443; IQR = 0.437–0.449), followed by nrITS2 (median = 0.527; IQR = 0.513–0.566), and *rbcL* (median = 0.584; IQR = 0.577–0.590; Figure 2). The range of nrITS2 linguistic complexity is relatively broad especially in comparison to that of *matK* and *rbcL*—perhaps a result of different functional constraints on structural versus protein coding sequences.

One might expect that sequences with lower linguistic complexity (i.e. those with homopolymers and/or simple sequence repeats) will have lower sequence quality due to slip-strand mispairing at the site of repetitive sequence elements [64,91–93], however lower linguistic complexity is correlated with higher sequence quality in Podocarpaceae ( $p < 2.2 \times 10^{-16}$ ). The sequences with the lowest linguistic complexity generally have sequence quality typical for the marker in question (Figure 2).

A homopolymer eight bases or longer was found in 35 sequences of the complete sample: 34 were *matK* sequences and one was an nrITS2 sequence. In the *matK* sequences from the complete sample, there is a single occurrence of A<sub>8</sub> and 33 occurrences of T<sub>8</sub>. The T<sub>8</sub> homopolymers occupy alignment positions 465–472 (found in all samples of *Afrocarpus*, *Lepidothamnus*, *Nageia*, *Prumnopitys*, and *Retrophyllum*) and 720–727 (found in some *Podocarpus*). The frequency of homopolymer occurrence was significantly different among markers ( $p=0.05$ ). Counter to previous findings [92,93], high homopolymer frequency is correlated with high sequence quality in Podocarpaceae ( $p = 2.1 \times 10^{-16}$ ). Previous investigations of the relationship between homopolymers and sequence quality focused on homopolymers ten bases or longer because they consistently result in low sequence quality [91–93], however homopolymers ten bases or longer are not found in any sequence of the complete sample. Thus we cannot determine if this length homopolymer has any effect on sequence quality.

The observed correlation between increased sequence quality and decreased linguistic complexity as well as the correlation between increased sequence quality and increased homopolymer frequency indicate that a mechanism other than slip-strand mispairing is responsible for the low quality sequences in the complete sample.

### Species discrimination

For individual markers, BRONX [74,75] species discrimination ranged from 28.8% to 38.1% (Figure 3; Table 1). Discrimination for marker combinations was slightly better at 46.4% to 56.7%. Discriminatory power did not statistically differ ( $p=0.05$ ) among markers or marker combinations. When species identification failed, the number of conflated species ranged from a mean of 4.3

( $\sigma = 1.8$ ) to 5.6 ( $\sigma = 4.1$ ) species for individual markers and a mean of 2.9 ( $\sigma = 1.4$ ) to 3.6 ( $\sigma = 2.1$ ) species for marker combinations. There were no unambiguous statistical differences ( $p=0.05$ ) in the number of conflated species among markers or marker combinations.

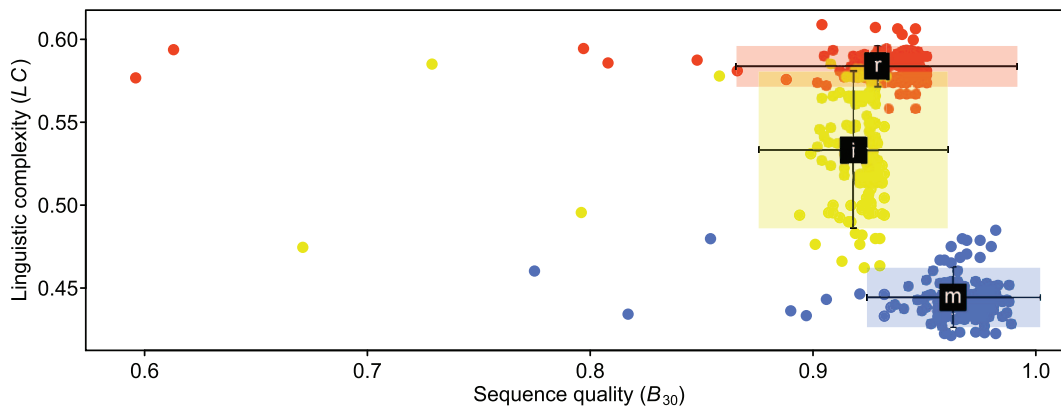
A synergistic effect was observed for marker combinations both in terms of an increase in discriminatory power and decrease in the number of conflated species (Table 1; e.g. the two specimens of *Af. mannii* examined cannot be consistently distinguished from *Af. dawei*, *Af. falcatus*, *Af. gracilior*, or *Af. usambarensis* by any single marker, but when *matK* and nrITS2 are combined, *Af. mannii* can be consistently distinguished from all other species). In no case did combining markers result in a loss of discriminatory power.

The core barcode markers (*matK* and *rbcL*) were able to consistently distinguish among 46.3% of the species in the complete sample (Figure 3). In comparison, studies that analyzed sequences of *matK*, *rbcL*, and nrITS2, individually and in combination, using comparable methods of species discrimination (the 'best match' procedure [94] or the 'simple pairwise matching' technique [95]) had a median success rate of 59.5% (range = 35.7–71.4) for core barcode markers (Table 2; [89,90,96–98]). In these same studies, species discrimination noticeably improved with the addition nrITS2 as a supplemental marker (median = 92.6%; range = 57.1–99.3). Although species discrimination did improve in Podocarpaceae with the addition of nrITS2 (Figure 3), the rate of species discrimination (56.7%) is less than the lowest published value (Table 2; [97]).

Of the 49 species represented by two or more individuals in the complete sample, BRONX could distinguish 28 (57.1%) from all other species using a combination of three markers (Table 1). In contrast, the 'least inclusive clade' method could distinguish 21 (42.9%) species (Figure 1). This provides another example of the poor performance of tree-based algorithms for barcode sequence discrimination [73,74].

The complete sample was composed of 97 species represented by 90 distinct multilocus genotypes. Thus, if intraspecific variation is assumed to be near zero, one could plausibly expect that species discrimination would be close to 92.8%, however only 56.7% of species could be consistently distinguished using all three markers simultaneously.

In many cases, identification failed in spite of ostensibly useful variation being present—this most often occurred when genotypes were shared among species (e.g. *Po. guatemalensis* and *Po. matudae* are sister species [54] that have a total of three multilocus genotypes [Figure 1]: the first multilocus genotype is restricted to *Po. guatemalensis*, the second multilocus genotype is restricted to *Po. matudae*, and the third multilocus genotype is found in both species). In the cases where genotypes are shared across species boundaries, the data cannot definitively distinguish between the underlying causal mechanisms of recent introgression versus



**Figure 2. Barcode sequence quality ( $B_{30}$ ) versus linguistic complexity (LC) for complete samples.** Circles represent individual *matK* (blue; m), *rbCL* (red; r), and nrITS2 (yellow; i) sequences. Black squares indicate marker means. Error bars span three standard deviations. doi:10.1371/journal.pone.0081008.g002

ancestral polymorphism followed by incomplete lineage sorting. In these cases, it is unlikely that sequence data from additional markers will increase species discrimination.

In some cases, identification failure is the result of an absence of sequence variation (e.g. *Dc. compactus* and *Dc. expansus* are sister species [54] that have identical sequences for all three markers [Figure 1]). Sequence data from additional markers may improve species discrimination in these cases. Although we did not test the utility of supplementary plastid markers, it seems unlikely that better discrimination will be provided by additional plastid data given the small difference in species discrimination between *matK* and *rbCL* (4.1%; Figure 3)—discriminatory power for plastid markers usually plateaus at two markers [95]. Rather than sequencing more plastid markers, effort would be better invested in variable unlinked markers that are easily recovered from Podocarpaceae (e.g. *NEEDLY* intron 2 [54,99]).

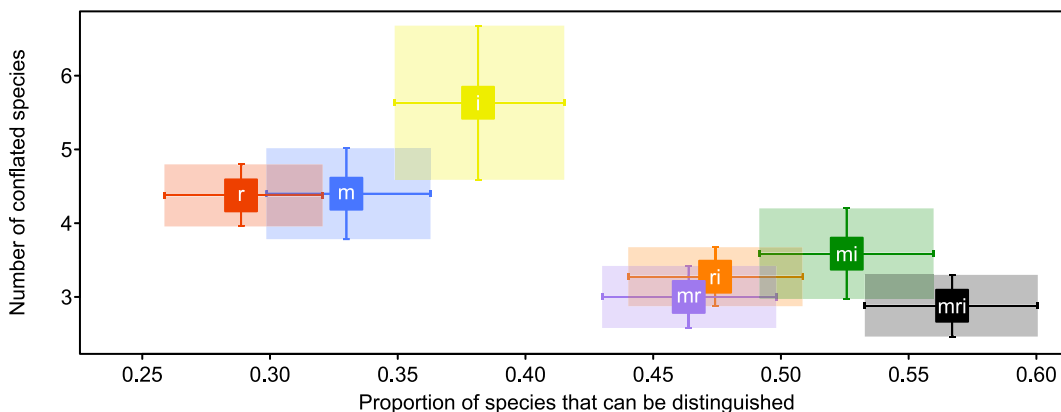
Discrimination success was mixed for the two CITES-listed Podocarpaceae species (Table 1): *Po. parlatoresi* (CITES Appendix I) can be distinguished from all other species using nrITS2 (*matK* and *rbCL* cannot distinguish *Po. parlatoresi* from *Po. sprucei*; *rbCL* cannot distinguish *Po. parlatoresi* from *Po. transiens*); *Po. nerifolius* (CITES Appendix III) cannot be distinguished from *Po. thailandensis* using all three markers (using single markers, *Po. nerifolius* can also be conflated with *Po. archboldii*, *Po. assamica*, *Po. brassii*, *Po. crassigemmis*,

*Po. drouynianus*, *Po. gibbsiae*, *Po. insularis*, *Po. ledermannii*, *Po. philippinensis*, *Po. polystachyus*, *Po. ramosii*, *Po. rubens*, and/or *Po. subtropicalis*). The herbal dietary supplement, *N. nagi* (Asian bayberry), cannot be distinguished from *N. formosensis* using all three markers (*matK* also cannot distinguish *N. nagi* from *N. motleyi*).

### Barcode gap

The barcode gap is a measure of the relative variation within and among species [78]. In the complete sample, 39.1% of species had a barcode gap for *matK*, 34.0% for *rbCL*, 38.1% for nrITS2, and 50.5% for all markers simultaneously (Figure 4; Table 1). There is no statistical difference ( $p=0.05$ ) in the frequency of barcode gaps among markers. The presence of a barcode gap is not correlated with sample size in Podocarpaceae ( $r_{pb}=0.06$ ;  $p=0.27$ ).

Barcode gaps quantify species distinctness at the barcode locus and thereby provide a crude measure of identification reliability (i.e. a species without a barcode gap may be more likely to be misidentified since it is not particularly distinctive; [71]). In this data set, whether a species can be consistently distinguished from all other species is unrelated to the presence or absence of a barcode gap ( $p=0.02$ ). For *matK* and *rbCL*, all of the species that can be consistently diagnosed have a barcode gap, but there are six



**Figure 3. Species discrimination by barcode marker for complete samples.** Squares indicate means for *matK* (blue; m), *rbCL* (red; r), nrITS2 (yellow; i), *matK* combined with *rbCL* (purple; mr), *matK* combined with nrITS2 (green; mi), *rbCL* combined with nrITS2 (orange; ri), and all markers combined (black; mri). Error bars indicate 95% confidence intervals. doi:10.1371/journal.pone.0081008.g003



**Table 2.** Rates of discriminatory success for barcoding studies that analyzed *matK*, *rbcl*, and nrITS2 sequences using algorithms comparable to BRONX.

study focus	<i>matK+rbcl</i>	<i>matK+rbcl+nrITS2</i>
<i>Parnassia</i> (Parnassiaceae) [98]	71.42%	96.19%
Chinese <i>Primula</i> sect. <i>Proliferae</i> (Primulaceae) [96]	68.75%	87.50%
Malagasy <i>Euphorbia</i> (Euphorbiaceae) [90]	<b>59.45%</b>	99.32%
Caryoteae (Arecaceae) [89]	51.85%	<b>92.59%</b>
<i>Actaea</i> (Ranunculaceae) [97]	35.71%	57.14%

Median rates of discriminatory success are in boldface.  
doi:10.1371/journal.pone.0081008.t002

species with barcode gaps that cannot be consistently differentiated from all other species (*matK*: *Dd. beccarii*, *Po. bracteatus*, *Po. novae-caledoniae*, *Po. nubigenus*, *Po. rumphii*, and *R. minus*; *rbcl*: *Manoao colensoi*, *N. wallichiana*, *Po. bracteatus*, *Po. pilgeri*, *Po. spinolosus*, and *Pr. andina*). In contrast, there are four species that do not have nrITS2 barcode gaps, but can be consistently diagnosed with nrITS2 (*Af. gracilior*, *Dc. imbricatus*, *Po. lambertii*, and *Pr. ferruginoides*). There are also four species that have nrITS2 barcode gaps that cannot be consistently differentiated from all other species using nrITS2 (*Po. bracteatus*, *Po. celatus*, *Po. longifoliolatus*, and *Po. sprucei*). There are no species with multilocus barcode gaps that cannot be consistently diagnosed using all three markers simultaneously, but there are six species that do not have multilocus barcode gaps that can be consistently diagnosed (*Af. gracilior*, *Af. mannii*, *Dc. imbricatus*, *Dc. kinabaluensis*, *Po. polystachyus*, and *Pr. ferruginoides*; Table 1).

The absence of a barcode gap coupled with discrimination success serves to contrast algorithmic approaches that use diagnostic nucleotide positions (i.e. those positions that consistently distinguish one species from all others) with distance-based methods. The presence of a barcode gap, does not guarantee that a species will be distinct. For example, a species may have a large amount of intraspecific variation combined with a small, but

consistent, amount of interspecific variation rendering the species without a barcode gap, but consistently diagnosable—one nucleotide difference that consistently differentiates the species in question from all other species is all that is required. Thus, the absence of a barcode gap is a poor predictor of discrimination success.

The presence of a barcode gap coupled with discrimination failure is an artifact of the analysis conducted: barcode gaps were computed using only sequences in the complete sample whereas discrimination was calculated with a reference database composed of all sequences. Thus, the barcode gap calculation did not necessarily include samples with zero interspecific distance that were included in the discrimination calculation. Restricting the discrimination calculation to sequences in the complete sample would have overestimated discrimination success for Podocarpaceae. At the same time, calculating the barcode gap using all sequences would have resulted in incomparable values.

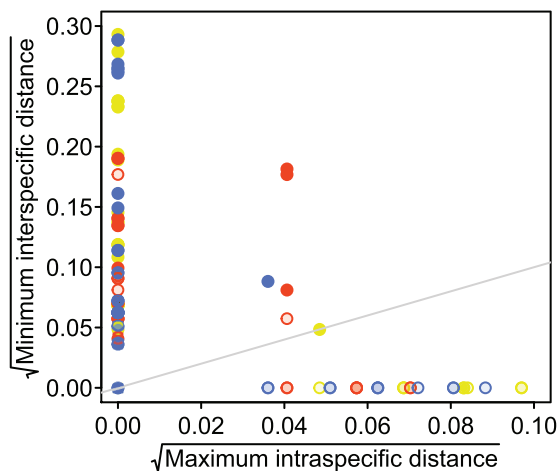
Sampling of additional individuals cannot decrease the maximum intraspecific distance, nor can it increase the minimum interspecific distance. Thus, new sequence data for *matK*, *rbcl*, and nrITS2 will either maintain or decrease the number of species with barcode gaps. Likewise, the rate of species discrimination cannot improve, and will most likely deteriorate, with additional sampling of individuals. New sequences of unlinked markers may however increase the number of species with barcode gaps and/or improve the rate of species discrimination.

## Conclusions

The vast majority of barcode sequences generated for this study were of high quality (Figure 2). Even the lowest quality sequences exceeded the minimum requirements of the BARCODE data standard. In the few instances that low quality sequences were generated, the responsible mechanism could not be discerned: slip-strand mispairing at the site of repetitive sequence elements cannot adequately explain the low quality sequences observed.

The power of *matK*, *rbcl*, and nrITS2, individually and in combination, to discriminate among Podocarpaceae species is relatively low (56.7% of species at maximum; Table 1; Figure 3). There were no statistically significant differences in the discriminatory power of markers or marker combinations. Although the discrimination rate for Podocarpaceae is below the rate reported for comparably analyzed studies (Table 2), it is not markedly lower. Plant DNA barcoding studies that heavily sample within taxonomic groups usually report low rates of species discrimination.

Discrimination success was mixed for Podocarpaceae species important in commerce and of conservation concern (Table 1). The CITES Appendix I species, *Po. parlatorei*, can be distinguished from all other species using nrITS2. Unfortunately, the CITES



**Figure 4. Barcode variation within and among species for complete samples.** Circles represent the set of *matK* (blue), *rbcl* (red), and nrITS2 (yellow) sequences for each species. Opaque filled circles denote diagnostic sequence sets. Non-diagnostic sequence sets are indicated with semi-transparent filled circles. Equal intra- and interspecific variation is marked by the gray line. Points above the gray line indicate species with 'barcode gaps'.  
doi:10.1371/journal.pone.0081008.g004

Appendix III species, *Po. neriiifolius*, and the herbal dietary supplement, *N. nagi*, cannot be unambiguously distinguished from all other Podocarpaceae using all three markers.

The presence of a barcode gap was not predictive of discrimination success. There was no statistically significant difference in the frequency of barcode gaps among markers in Podocarpaceae (Figure 4). In addition, there was no correlation between number of individuals sampled per species and the presence of a barcode gap.

Sequences of additional variable unlinked markers that are easily recovered from Podocarpaceae (e.g. *NEEDLY* intron 2) may increase the rate of species discrimination.

## References

- Farjon A (2001) World checklist and bibliography of conifers. Richmond: Royal Botanic Gardens, Kew, 2 edition.
- Cernusak LA, Adie H, Bellingham PJ, Biffin E, Brodribb TJ, et al. (2011) Podocarpaceae in tropical forests: a synthesis. In: Turner BL, Cernusak LA, editors, Ecology of the Podocarpaceae in tropical forests, Washington, D.C.: Smithsonian Institution Scholarly Press, volume 95. pp. 189–195.
- de Laubenfels DJ (1969) A revision of the Malesian and pacific rainforest conifers, i. Podocarpaceae, in part. *Journal of the Arnold Arboretum* 50: 315–369.
- de Laubenfels DJ (1988) *Coniferales*. *Flora Malesiana* 10: 337–453.
- Buchholz JT, Gray NE (1948) A taxonomic revision of *Podocarpus* I: the sections of the genus and their subdivisions with special reference to leaf anatomy. *Journal of the Arnold Arboretum* 29: 46–63.
- Buchholz JT, Gray NE (1948) A taxonomic revision of *Podocarpus* II: the American species of *Podocarpus* section *Stachycarpus*. *Journal of the Arnold Arboretum* 29: 64–76.
- Buchholz JT, Gray NE (1948) A taxonomic revision of *Podocarpus* IV: the American species of section *Eupodocarpus* subsections C and D. *Journal of the Arnold Arboretum* 29: 123–151.
- Gray NE, Buchholz JT (1948) A taxonomic revision of *Podocarpus* III: the American species of *Podocarpus* section *Polypodioopsis*. *Journal of the Arnold Arboretum* 29: 117–122.
- Gray NE, Buchholz JT (1951) A taxonomic revision of *Podocarpus* V: the south Pacific species of *Podocarpus* section *Stachycarpus*. *Journal of the Arnold Arboretum* 32: 82–92.
- Gray NE, Buchholz JT (1951) A taxonomic revision of *Podocarpus* VI: the south Pacific species of *Podocarpus* section *Sundacarpus*. *Journal of the Arnold Arboretum* 32: 93–98.
- Gray NE (1953) A taxonomic revision of *Podocarpus* VII: the African species of *Podocarpus* section *Afrocarpus*. *Journal of the Arnold Arboretum* 34: 67–76.
- Gray NE (1953) A taxonomic revision of *Podocarpus* VIII: the African species of section *Eupodocarpus*, subsections A and E. *Journal of the Arnold Arboretum* 34: 163–175.
- Gray NE (1955) A taxonomic revision of *Podocarpus* IX: the south Pacific species of section *Eupodocarpus*, subsection F. *Journal of the Arnold Arboretum* 36: 199–206.
- Gray NE (1956) A taxonomic revision of *Podocarpus* X. *Journal of the Arnold Arboretum* 37: 160–172.
- Gray NE (1958) A taxonomic revision of *Podocarpus* XI. *Journal of the Arnold Arboretum* 39: 424–477.
- Gray NE (1960) A taxonomic revision of *Podocarpus* XII: section *Microcarpus*. *Journal of the Arnold Arboretum* 41: 36–39.
- Gray NE (1962) A taxonomic revision of *Podocarpus* XIII: section *Polypodioopsis* in the south Pacific. *Journal of the Arnold Arboretum* 43: 67–79.
- Schoonraad E, Vanderschijff HP (1974) Anatomy of leaves of genus *Podocarpus* in South Africa. *Phytomorphology* 24: 75–85.
- Knopf P, Nimsh H, Stützel T (2007) *Dacrydium* × *suprinii*, sp. Nova—a natural hybrid of *Dacrydium araucarioides* × *D. guillauminii*. *Feddes Repertorium* 118: 51–59.
- Knopf P (2011) Differential diagnosis and evolution within the Podocarpaceae s. l. Ph.D. thesis, Ruhr–University Bochum.
- Stockey RA, Ko H (1990) Cuticle micromorphology of *Dacrydium* (Podocarpaceae) from New Caledonia. *Botanical Gazette* 151: 138–149.
- Stockey RA, Ko H, Woltz P (1992) Cuticle micromorphology of *Falcatifolium* de Laubenfels (Podocarpaceae). *International Journal of Plant Sciences* 153: 589–601.
- Stockey RA, Ko H, Woltz P (1995) Cuticle micromorphology of *Parasitaxus* de Laubenfels (Podocarpaceae). *International Journal of Plant Sciences* 156: 723–730.
- Stockey RA, Frevel BJ (1997) Cuticle micromorphology of *Prumnopitys Philippi* (Podocarpaceae). *International Journal of Plant Sciences* 158: 198–221.
- Stockey RA, Frevel BJ, Woltz P (1998) Cuticle micromorphology of *Podocarpus*, subgenus *Podocarpus*, section *Scytopodium* (Podocarpaceae) of Madagascar and South Africa. *International Journal of Plant Sciences* 159: 923–940.
- Mill RR, Schilling DM (2009) Cuticle micromorphology of *Saxegothaea* (Podocarpaceae). *Botanical Journal of the Linnean Society* 159: 58–67.
- Schilling DMS, Mill RR (2011) Cuticle micromorphology of Caribbean and Central American species of *Podocarpus* (Podocarpaceae). *International Journal of Plant Sciences* 172: 601–631.
- IUCN (2012) International union for the conservation of nature red list of threatened species, version 2012.2. Available: <http://www.iucnredlist.org>. Accessed 2013 March 5.
- CITES (2012) Convention on international trade in endangered species, appendices I, II, and III. Available: <http://www.cites.org>. Accessed 2013 March 5.
- McGuffin M, Kartesz JT, Leung AY, Tucker AO (2000) Herbs of Commerce. American Herbal Products Association, 2nd edition.
- Fu L, Li Y, Mill RR (1994) Podocarpaceae. In: Wu Z, Raven PH, Hong D, editors, *Flora of China*, St. Louis: Missouri Botanical Garden, volume 4. pp. 78–84.
- Facciola S (1990) *Cornucopia: a source book of edible plants*. Vista: Kampong Publications.
- Abdillahi HS, Stafford GI, Finnie JF, Van Staden J (2010) Ethnobotany, phytochemistry and pharmacology of *Podocarpus* sensu latissimo (s.l.). *South African Journal of Botany* 76: 1–24.
- Abdillahi HS, Verschaeve L, Finnie JF, Van Staden J (2012) Mutagenicity, antimutagenicity and cytotoxicity evaluation of South African *Podocarpus* species. *Journal of Ethnopharmacology* 139: 728–738.
- Bauch J, Schmidt O, Hillis WE, Yazaki Y (1977) Deposits in heartshakes of *Dacrydium* species and their toxicity against fungi and bacteria. *Holzforschung* 31: 1–7.
- Symonds EL, Konczak I, Fenech M (2012) The Australian fruit illawarra plum (*Podocarpus elatus* Endl., Podocarpaceae) inhibits telomerase, increases histone deacetylase activity and decreases proliferation of colon cancer cells. *British Journal of Nutrition* 15: 1–9.
- Abdillahi HS, Finnie JF, Van Staden V (2008) Antibacterial activity of *Podocarpus* species. *South African Journal of Botany* 74: 359–360.
- Abdillahi HS, Stafford GI, Finnie JF, Van Staden J (2008) Antimicrobial activity of South African *Podocarpus* species. *Journal of Ethnopharmacology* 119: 191–194.
- Hayashi Y, Matsumoto T, Tashiro T (1979) Antitumor activity of norditerpenoid dilactones in *Podocarpus* plants—structure activity relationship on in vitro cytotoxicity against Yoshida sarcoma. *Gann* 70: 365–369.
- Hembree JA, Chang CJ, McLaughlin JL, Cassady JM, Watts DJ, et al. (1979) Cytotoxic norditerpenoid dilactones of *Podocarpus milanjianus* and *Podocarpus seelowii*. *Phytochemistry* 18: 1691–1694.
- Hembree JA, Chang C, McLaughlin JL, Cassady J (1980) Milanjilactone A and milanjilactone B, 2 novel cytotoxic norditerpenoid dilactones from *Podocarpus milanjianus* Rendle. *Experientia* 36: 28–29.
- Park HS, Takahashi Y, Fukaya H, Aoyagi Y, Takeya K (2003) S-R-podolactone D, a new sulfoxide-containing norditerpenoid dilactone from *Podocarpus macrophyllum* var. *maki*. *Journal of Natural Products* 66: 282–284.
- Park HS, Takahashi Y, Fukaya H, Aoyagi Y, Takeya K (2004) New cytotoxic norditerpenoid dilactones from leaves of *Podocarpus macrophyllum* var. *maki*. *Heterocycles* 63: 347–357.
- Bergin DO (2000) Current knowledge relevant to management of *Podocarpus totara* for timber. *New Zealand Journal of Botany* 38: 343–359.
- Phillips EWJ (1941) The identification of coniferous woods by their microscopic structure. *The Journal of the Linnean Society of London, Botany* 52: 259–320.
- Dallimore W, Jackson AB, Harrison SG (1967) *A handbook of Coniferae and Ginkgoaceae*. New York: St. Martin's Press.
- Farjon A (2008) *A natural history of conifers*. Portland: Timber Press.

## Supporting Information

**Dataset S1** Vouchers and GenBank accessions for samples of Podocarpaceae. (TAB)

## Acknowledgments

We thank Joan Deutsch for providing excellent technical assistance. License for the use of TNT was generously provided by the Willi Hennig Society.

## Author Contributions

Conceived and designed the experiments: DPL PK CS. Performed the experiments: PK CS. Analyzed the data: DPL CS. Contributed reagents/materials/analysis tools: DPL PK CS. Wrote the paper: DPL PK CS.

48. Cockayne L (1919) New Zealand plants and their story. Wellington: M. F. Marks, Government Printer.
49. Kubo I, Matsumoto T, Klocke JA (1984) Multichemical resistance of the conifer *Podocarpus gracilior* (Podocarpaceae) to insect attack. *Journal of Chemical Ecology* 10: 547–559.
50. Russell GB, Singh P, Fenemore PG (1972) Insect-control chemicals from plants: naglactone c, a toxic substance from the leaves of *Podocarpus nivalis* and *P. hallii*. *Australian Journal of Biological Sciences* 25: 1025–1029.
51. Saeki I, Sumimoto M, Kondo T (1970) Termiticidal substances from wood of *Podocarpus macrophyllus* D. Don. *Holzforschung* 24: 83–86.
52. Li Y, Gao LM, Poudel RC, Li DZ, Forrest A (2011) High universality of *matK* primers for barcoding gymnosperms. *Journal of Systematics and Evolution* 49: 169–175.
53. Chiou SJ, Yen JH, Fang CL, Chen HL, Lin TY (2007) Authentication of medicinal herbs using PCR-amplified ITS2 with specific primers. *Planta Medica* 73: 1421–1426.
54. Knopf P, Schulz C, Little DP, Stützel T, Stevenson DW (2012) Relationships within Podocarpaceae based on DNA sequence, anatomical, morphological, and biogeographical data. *Cladistics* 28: 271–299.
55. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.
56. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.
57. Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analysis. *Systematic Biology* 49: 369–381.
58. Little DP (2005) 2xread: a simple indel coding tool. Available: <http://www.nybg.org/files/scientists/2xread.html>. Accessed 2013 January 5.
59. Farris JS, Albert VA, Källersjö M, Lipscomb D, Kluge AG (1996) Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12: 99–124.
60. Goloboff PA, Farris JS, Nixon KC (2008) TNT, a free program for phylogenetic analysis. *Cladistics* 24: 774–786.
61. Little DP (2010) A unified index of sequence quality and contig overlap for DNA barcoding. *Bioinformatics* 26: 2780–2781.
62. Pesole G, Attimonelli M, Saccone C (1996) Linguistic analysis of nucleotide sequences: algorithms for pattern recognition and analysis of codon strategy. *Methods in Enzymology* 266: 281–294.
63. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics* 16: 276–277.
64. Shinde D, Lai Y, Sun F, Arnheim N (2003) *Taq* DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)*n* and (A/T)*n* microsatellites. *Nucleic Acids Research* 31: 974–980.
65. Scheffé H (1953) A method for judging all contrasts in the analysis of variance. *Biometrika* 40: 87–104.
66. de Mendiburu F (2012) *Agricolae* version 1.1-3. Available: <http://cran.r-project.org/>. Accessed 2013 January 5.
67. R development core team (2012) R: a language and environment for statistical computing (version 2.15.2). Vienna, Austria: R Foundation for Statistical Computing.
68. Spearman C (1904) The proof and measurement of association between two things. *The American Journal of Psychology* 15: 72–101.
69. Salinas NR, Little DP (2013) 2matrix: a utility for indel coding and phylogenetic matrix concatenation. Available: <https://github.com/nrsalinas/2matrix>. Accessed 2013 September 13.
70. Nixon KC (1999) The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics* 15: 407–414.
71. Sukumaran J, Holder MT (2010) DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26: 1569–1571.
72. Paradis E, Bolker B, Claude J, Cuong HS, Desper R, et al. (2013) The APE (Analyses of Phylogenetics and Evolution) package version 3.0-10. Available: <http://cran.r-project.org/>. Accessed 2013 January 5.
73. Little DP, Stevenson DW (2007) A comparison of algorithms for identification of specimens using DNA barcodes: examples from gymnosperms. *Cladistics* 23: 1–21.
74. Little DP (2011) DNA barcode sequence identification incorporating taxonomic hierarchy and within taxon variability. *PLoS ONE* 6: e20552.
75. Little DP (2012) BRONX: Barcode Recognition Obtained with Nucleotide eXposés version 2.0. Available: <http://www.nybg.org/files/scientists/dlitle/BRONX2.html>. Accessed 2013 January 5.
76. Wilson EB (1927) Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association* 22: 209–212.
77. Harrell Jr FE (2012) The Hmisc package version 3.10-1. Available: <http://cran.r-project.org/>. Accessed 2013 January 5.
78. Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology* 3: e422.
79. Levenshtein VI (1966) Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics|Doklady* 10: 707–710.
80. Meier R, Zhang G, Ali F (2008) The use of mean instead of smallest interspecific distances exaggerates the size of the “barcoding gap” and leads to misidentification. *Systematic Biology* 57: 809–813.
81. Rizopoulos D (2013) Latent Trait Models under IRT version 0.9-9. Available: <http://cran.r-project.org/>. Accessed 2013 January 5.
82. McNemar Q (1947) Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12: 153–157.
83. Conran JG, Wood GM, Martin PG, Dowd JM, Quinn CJ, et al. (2000) Generic relationships within and between the gymnosperm families Podocarpaceae and Phyllocladaceae based on an analysis of the chloroplast gene *rbcL*. *Australian Journal of Botany* 48: 715–724.
84. Kelch DG (2002) Phylogenetic assessment of the monotypic genera *Sundacarpus* and *Manoao* (Coniferales: Podocarpaceae) utilising evidence from 18S rDNA sequences. *Australian Systematic Botany* 15: 29–35.
85. Sinclair WT, Mill RR, Gardner MF, Woltz P, Jaffré T, et al. (2002) Evolutionary relationships of the New Caledonian heterotrophic conifer *Parasitaxus usta* (Podocarpaceae), inferred from chloroplast *trnL-F* intron/spacer and nuclear rDNA ITS2 sequences. *Plant Systematics and Evolution* 233: 79–104.
86. Biffin E, Conran JC, Lowe AJ (2011) Podocarp evolution: a molecular phylogenetic perspective. In: Turner BL, Cernusak LA, editors, *Ecology of the Podocarpaceae in tropical forests*, Washington, D.C.: Smithsonian Institution Scholarly Press, volume 95. pp. 1–20.
87. Farris JS (1974) Formal definitions of paraphyly and polyphyly. *Systematic Zoology* 23: 548–554.
88. Hanner R (2009) Proposed standards for BARCODE records in INSDC (BRIs). Technical report, Database Working Group, Consortium for the Barcode of Life. Available: [http://barcoding.si.edu/PDF/DWG\\_data\\_standards-Final.pdf](http://barcoding.si.edu/PDF/DWG_data_standards-Final.pdf). Accessed 2013 January 5.
89. Jeanson ML, Labat JN, Little DP (2011) DNA barcoding: a new tool for palm taxonomists? *Annals of Botany* 108: 1445–1451.
90. Aubriot X, Lowry PP, Cruaud C, Couloux A, Haevermans T (2013) DNA barcoding in a biodiversity hot spot: potential value for the identification of *Malagasy Euphorbia L.* listed in CITES Appendices I and II. *Molecular Ecology Resources* 13: 57–65.
91. Devey DS, Chase MW, Clarkson JJ (2009) A stuttering start to plant DNA barcoding: microsatellites present a previously overlooked problem in non-coding plastid regions. *Taxon* 58: 7–15.
92. Fazekas AJ, Steeves R, Newmaster SG (2010) Improving sequencing quality from PCR products containing long mononucleotide repeats. *BioTechniques* 48: 277–285.
93. Fazekas AJ, Steeves R, Newmaster SG, Hollingsworth PM (2010) Stopping the stutter: improvements in sequence quality from regions with mononucleotide repeats can increase the usefulness of non-coding regions for DNA barcoding. *Taxon* 59: 694–697.
94. Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tail of high intraspecific variability and low identification success. *Systematic Biology* 55: 715–728.
95. CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences* 106: 12794–12797.
96. Yan HF, Hao G, Hu CM, Ge XJ (2011) DNA barcoding in closely related species: a case study of *Primula L.* sect. *Polioferae* Pax (Primulaceae) in China. *Journal of Systematics and Evolution* 49: 225–236.
97. Baker DA, Stevenson DW, Little DP (2012) DNA barcode identification of black cohosh herbal dietary supplements. *Journal of AOAC International* 95: 1023–1034.
98. Yang JB, Wang YP, Möller M, Gao LM, Wu D (2012) Applying plant DNA barcodes to identify species of *Parnassia* (Parnassiaceae). *Molecular Ecology Resources* 12: 267–275.
99. Little DP (2004) Documentation of hybridization between Californian cypresses: *Cupressus macnabiana* × *sargentii*. *Systematic Botany* 29: 825–833.