Allopolyploid Speciation Accompanied by Gene Flow in a Tree Fern

Jing Wang,[1,2] Shiyong Dong,[1,2] Lihua Yang,[1,2] Aj Harris,[1,3] Harald Schneider,*[,4] and Ming Kang *[,1,2]

[1]Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

[2]Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Guangzhou, China

[3]Department of Biology, Oberlin College, Oberlin, OH

[4]Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, China

***Corresponding authors:** E-mails: mingkang@scbg.ac.cn; harald@xtbg.ac.cn.

Associate editor: Joanna Kelley

Raw illumina sequences of RAD sequences have been deposited at NCBI sequence read archive (accession no. PRJNA556085). Voucher specimens for newly sampled populations are deposited in the herbarium at South China Botanical Garden (IBSC).

Article

Abstract

Hybridization in plants may result in hybrid speciation or introgression and, thus, is now widely understood to be an important mechanism of species diversity on an evolutionary timescale. Hybridization is particularly common in ferns, as is polyploidy, which often results from hybrid crosses. Nevertheless, hybrid speciation as an evolutionary process in fern lineages remains poorly understood. Here, we employ flow cytometry, phylogeny, genomewide single nucleotide polymorphism data sets, and admixture and coalescent modeling to show that the scaly tree fern, *Gymnosphaera metteniana* is a naturally occurring allotetraploid species derived from hybridization between the diploids, *G. denticulata* and *G. gigantea*. Moreover, we detected ongoing gene flow between the hybrid species and its progenitors, and we found that *G. gigantea* and *G. metteniana* inhabit distinct niches, whereas climatic niches of *G. denticulata* and *G. metteniana* largely overlap. Taken together, these results suggest that either some degree of intrinsic genetic isolation between the hybrid species and its parental progenitors or ecological isolation over short distances may be playing an important role in the evolution of reproductive barriers. Historical climate change may have facilitated the origin of *G. metteniana*, with the timing of hybridization coinciding with a period of intensification of the East Asian monsoon during the Pliocene and Pleistocene periods in southern China. Our study of allotetraploid *G. metteniana* represents the first genomic-level documentation of hybrid speciation in scaly tree ferns and, thus, provides a new perspective on evolution in the lineage.

Key words: **allopolyploidy, coalescent, Cyatheaceae, Gymnosphaera, hybridization speciation, RADseq.**

Introduction

Polyploidization is considered to be an efficient mechanism leading to instant speciation due to the immediate establishment of strong reproductive barriers between the polyploid offspring and their ancestor(s) (Coyne and Orr 2004; Rieseberg and Willis 2007; Ramsey 2011). Two processes of polyploidization are considered to be distinct, namely autopolyploidy and allopolyploidy, and the later process comprises hybridization between two divergent species (Soltis PS and Soltis DE 2009). Typically, divergent species have reproductive barriers due to distinct cytotypes, which yield sterile triploid hybrid. However, gene flow between the offspring and diploid progenitors may be enabled through unreduced gametes (Ramsey and Schemske 1998), as demonstrated in several flowering plant species (Petit et al. 1999; Slotte et al. 2008; Ma et al. 2010; Zohren et al. 2016; Sutherland and Galloway 2017). This suggests that the conventional understanding of polyploid hybrids in complete isolation from their progenitors may not always hold true and that interploid gene flow may sometimes continue to occur. However, interploid gene flow has not been demonstrated to occur in ferns (Perrie et al. 2010), which have rates of polyploid speciation approximately twice that of angiosperms (Wood et al. 2009). Nevertheless, our understanding of the roles of hybridization and polyploidy in fern evolution speciation is limited in comparison with some lineages of angiosperms.

Hybridization is common in ferns especially due to their reproductive systems comprising both a free-living sporophyte and a free-living gametophyte. Through this reproductive system, high rates of gene flow is believed to occur among ferns, therefore, reducing the rate of formation of reproductive barriers and resulting in slower rates of speciation. The slower speciation rate may, in turn, represent a possible explanation for less species diversity in ferns compared with angiosperms (Ranker and Sundue 2015). However, this assumes that ferns have a great capacity for dispersal of spores or other propagules over geographic barriers that may otherwise intervene in gene flow (Chung MY and Chung MG 2013; Ramírez-Barahona and Eguiarte 2015). Nevertheless, natural hybrids appear to occur frequently between both

closely related and highly diverged fern taxa, and most of these hybridization events appear to yield polyploid offspring (e.g., Walker 1961; Wagner et al. 1992; Xiang et al. 2000; Grusz et al. 2009; Sigel 2016; Schneider et al. 2017). Broadly, it remains unclear whether hybridization in ferns has the evolutionary consequence of widespread, hyper-diverse species or frequent polyploid speciation.

Recent progress in the establishment of new genomic tools have enabled analysis of genomewide patterns of hybridization in a variety of systems (Payseur and Rieseberg 2016), and statistical models have been developed that utilize genomewide data for more powerful testing of specific evolutionary histories and the scenarios involved in allopolyploid speciation (Roux and Pannell 2015). The application of these tools has yielded considerable progress toward an understanding of polyploid speciation in angiosperms in recent years (Douglas et al. 2015; Vallejo-Marín et al. 2015; Luo et al. 2017). However, until now, the evolutionary mechanisms underlying polyploid speciation via hybridization in ferns has remained poorly understood using these new approaches in part due to a paucity of available genomic markers. In particular, no study has attempted to investigate the genomewide patterns of hybridization and gene flow among fern species.

Scaly tree ferns of the family Cyatheaceae are one of the most species rich families of ferns and comprise a major component of the order Cyatheales (PPG I 2016). This pantropical family of ferns comprises about 643 species and is usually characterized by an arborescent habit and pronounced local endemism (Korall and Pryer 2014). Phylogenetic studies have shown evidence for four major lineages with Cyatheaceae corresponding to the four recognized genera: *Alsophila* R. Br., *Cyathea* Smith, *Gymnosphaera* Blume, and *Sphaeropteris* Bernh (Korall et al. 2006; Janssen et al. 2008; Dong and Zuo 2018). All four genera occur in both United States and Australasia, and *Alsophila* and *Gymnosphaera* also have species in Africa (Korall and Pryer 2014). Unlike most ferns, which exhibit a fairly low degree of endemism due to high dispersal ability via wind-blown spores (Barrington 1993), Cyatheaceae has a large number of endemics occurring in subtropical montane forests (Janssen et al. 2008; Ramírez-Barahona et al. 2011). However, the cause for the high levels of species richness and endemism of the scaly tree ferns remains incompletely understood. Based on morphological evidence, the occurrence of natural interspecific hybrids has been considered in Cyatheaceae, but polyploidy in the family appears rare, suggesting that homoploid hybrid speciation may occur commonly (Conant 1975, 1990; Conant and Cooper-Driver 1980; Caluff 2002; Janssen and Rakotondrainibe 2007). In particular, in the study of natural hybridization among three species of *Alsophila*, Conant and Cooper-Driver (1980) proposed a model of autogamous allohomoploid speciation, in which new homoploid hybrid species can arise when allohomoploid F_1 hybrids are capable of gametophytic selfing and the recombinant F_2 individuals are at least partially reproductively isolated from their progenitor lineages. To date, most species in Cyatheaceae were reported to be diploid with diploid chromosome number $2n=138$ (Rice et al. 2015; Schneider, unpublished), which is seemingly consistent with the idea that homoploid hybrid speciation has been frequent in the family. Nevertheless, *Gymnosphaera metteniana* (Hance) Tagawa was reported as a tetraploid with $2n = 274$ (Nakato 1989), yet its evolutionary origins and mechanisms of speciation have not been explored.

Gymnosphaera is a small genus within the scaly tree ferns including ca. 30 species in Asia, 12 in Africa, and 2 (*G. capensis* [L. f.] S.Y. Dong and *G. salvinii* [Hook.] S.Y. Dong) in the Americas (Korall et al. 2006; Dong and Zuo 2018). It is the only genus of tree ferns where polyploidy has been reported and in which there is an intercontinental disjunction at the species level. The species *G. capensis* occurs in two continents, Africa and South America (Korall and Pryer 2014). Hybridization between species of *Gymnosphaera* has been suggested based on multilocus DNA sequence data (Dong et al. 2019). Here, we examine the evidence for hybridization and polyploidization in the evolutionary history of *Gymnosphaera*, focusing on three species mainly occurring in southern China and Indochina: *G. denticulata* (Baker) Copel., *G. gigantea* (Wall. ex Hook.) J. Sm., and *G. metteniana*. *Gymnosphaera denticulata* is the smallest species of scaly tree fern in mainland Asia characterized by the lack of trunks and having bullate scales on abaxial surface of leaf axes, whereas *G. gigantea* is much larger, with trunks up to 3.5 m tall and copious spreading scales throughout either side of stipes. In contrast, *G. metteniana* is a shrub-like scaly tree fern with short, usually decumbent trunks. Phytogeographically, *G. gigantea* is a tropical species occurring mainly in Indochina and Southwest China, whereas *G. denticulata* and *G. metteniana* are both distributed in subtropical areas of eastern Asia, namely in China and Japan (Zhang and Nishida 2013). The ranges of the latter two are generally overlapping, with *G. denticulata* extending north slightly more than *G. metterniana*. According to Dong et al. (2019), based on the analysis of nuclear sequences, *G. denticulata* and *G. gigantea* diverged in the late Neogene ca. 5.9 Ma. *Gymnosphaera metteniana* is morphologically intermediate between *G. denticulata* and *G. gigantea* (table 1), with which it is parapatric/sympatric and allopatric, respectively (fig. S1, Supplementary Material online). Preliminary phylogenetic analyses with chloroplast DNA (cpDNA) show that *G. metteniana* is nested within the *G. gigantea* clade (Dong and Zuo 2018), whereas multilocus nuclear DNA sequences revealed a sister relationship between *G. metteniana* and *G. denticulata* (Dong et al. 2019). Given tetraploidy in *G. metteniana* along with its positions within prior nuclear and plastid phylogenies, this species may have arisen through interspecies hybridization between *G. denticulata* and *G. gigantea*.

In this study, we combined analyses of cytology via flow cytometry, phylogeny, population genomics, and climatic niches to test the hypothesis that *G. metteniana* originated through hybridization between *G. denticulata* and *G. gigantea*. As the large size of genomes in Cyatheaceae hampers whole-genome sequencing (Clark et al. 2016), we employed a genome-reduction sequencing approach in which we obtained large sets of genome-wide single nucleotide polymorphisms (SNPs) to characterize patterns of genetic

Table 1. Comparison of Morphological Traits of *Gymnosphaera denticulata*, *G. gigantea*, and *G. metteniana*.

Character	*Gymnosphaera denticulata*	*Gymnosphaera metteniana*	*Gymnosphaera gigantea*
Trunk habit	Lacking	Decumbent, rarely erect	Erect
Trunk length	0	0–0.5 m	0.5–3.5 m
Lamina outline	Ovate-triangular	Ovate	Lanceolate
Position of scales on stipe	On lower part of stipe	On lower to middle or upper part of stipe	Throughout stipe
Stipe color	Light castaneous	Castaneous or blackish	Blackish
Arrangement of lower pinnae on rachis	Alternate	Alternate or sometimes subopposite	Opposite
Length of the petiolule of basal pinnae	1–4 cm	0.5–1 cm	0
Number of pinnae to either side of rachis	6 (5–7)	9 (6–12)	10 (7–16)
Scales on abaxial surface of costules	Bullate	Flat	Flat
Arrangement of sori on ultimate lobes	Two-rowed, parallel	Two-rowed, parallel or slightly V-shaped	V-shaped

Table 2. Estimated Ploidy Level, Mean Genome Size (2C), and SD of Genome Size for the Three Species of *Gymnosphaera*.

Species/Population	*N*	2C mean (pg) ± SD	Ploidy
Gymnosphaera denticulate	9	13.60±0.89	
Dong4718	3	14.56±0.95	2×
Dong4726	3	13.30±0.13	2×
Dong4738	3	12.95±0.30	2×
Gymnosphaera metteniana	15	28.84±1.56	
Dong4720	3	30.81±2.54	4×
Dong4725	3	28.55±0.20	4×
Dong4741	1	26.78	4×
Dong5106	3	29.03±0.54	4×
Dong4753	4	28.05±0.69	4×
SCBG02	1	28.54	4×
Gymnosphaera gigantea	10	17.28±0.39	
Dong5107	3	17.56±0.13	2×
Dong4743	2	17.42±0.17	2×
Dong4754	3	17.27±0.42	2×
SCBG01	2	16.74±0.34	2×

NOTE.—*N*, number of individuals analyzed for each population/species.

differentiation and hybridization in the three species of *Gymnosphaera*. We show that the tetraploid *G. metteniana* represents genomic intermediacy between the two proposed parental species. We then tested whether *G. metteniana* arose from multiple hybridization events. We also employed coalescent-based modeling to quantify gene flow between the hybrid and its progenitors after the polyploidization speciation that we detected. Finally, we used climatic data for each species to test whether the putative hybrid species occupies a distinct ecological niche with respect to the parents, which would suggest that reproductive isolation could be achieved via geographic isolation.

Results

Genome Size and Ploidy

We determined the relative DNA content of the three species, based on samples from 11 natural populations (with 1–4 individuals per population) and three living collections in cultivation at South China Botanical Garden (SCBG). Mean 2 C-values were 17.28, 13.60, and 28.84 pg for *G. gigantea*, *G. denticulata*, and *G. metteniana*, respectively (table 2 and supplementary table S1, Supplementary Material online). Evidence for intraspecific genome size variation was detected, and the 2 C-values differed by 6.7%, 19.4%, and 20.5% for *G. gigantea*, *G. denticulata*, and *G. metteniana*, respectively. Despite the apparent intraspecific genome size variation, the results revealed up to about 2-fold 2 C DNA amount difference between *G. metteniana* and the other two species, consistent with tetraploidy in *G. metteniana*. The results from ploidy analysis were consistent with the genome size estimates, further supporting that *G. metteniana* is tetraploid.

Phylogenetic Evidence of Hybridization

We tested the hypothesis of allopolyploid speciation of *G. metteniana* and aimed to determine the maternal and paternal progenitors by sequencing five cpDNA markers and nine nuclear single-copy nuclear DNA (nDNA). The cpDNA phylogeny (4,157 bp; supplementary table S2, Supplementary Material online) revealed that all individuals of *G. metteniana*, *G. gigantea*, and *G. andersonii* (J. Scott ex Bedd.) Ching and S.K. Wu formed a clade (Bootstraps [BS] = 100%), which was found to be the sister group (BS = 50%) to a clade (BS = 100%) containing all samples of *G. austroyunnanensis* (S.G. Lu) S.G. Lu and C.X. Li, *G. salletii* (Tardieu & C. Chr.) S.Y. Dong, and one species new to science (fig. 1*A* and supplementary fig. S2, Supplementary Material online). *Gymnosphaera denticulata* and *G. podophylla* were found to be monophyletic (BS = 100%) by the cpDNA, and they were inferred as sister species. Our phylogenetic analyses of nDNA (5,470 bp; supplementary table S3, Supplementary Material online) demonstrated that two homologs present in most *G. metteniana* individuals (37 out of 43) grouped with one each of the two putative parents (fig. 1*B* and supplementary fig. S3, Supplementary Material online). This pattern of discordance between the nuclear and plastid data provides strong support for the hypothesis that tetraploid *G. metteniana* originated via hybridization between *G. gigantea*/*G. andersonii* and *G. denticulata*. Given that cpDNA is predominantly maternally inherited in ferns (Gastony and Yatskievych 1992), our results further suggest that either *G. gigantea* or *G. andersonii* likely served as the maternal parent in the initial hybridization events. However, it is unlikely that *G. andersonii* was the maternal donor, because it is a geographically isolated, endemic species in southwest China and southern Asia (Dong and Zuo 2018). In addition, *G. andersonii* is very different morphologically from all other *Gymnosphaera* species by having fronds abaxially covered with hairs (Dong and Zuo 2018). The morphological evidence taken together with the result of cpDNA

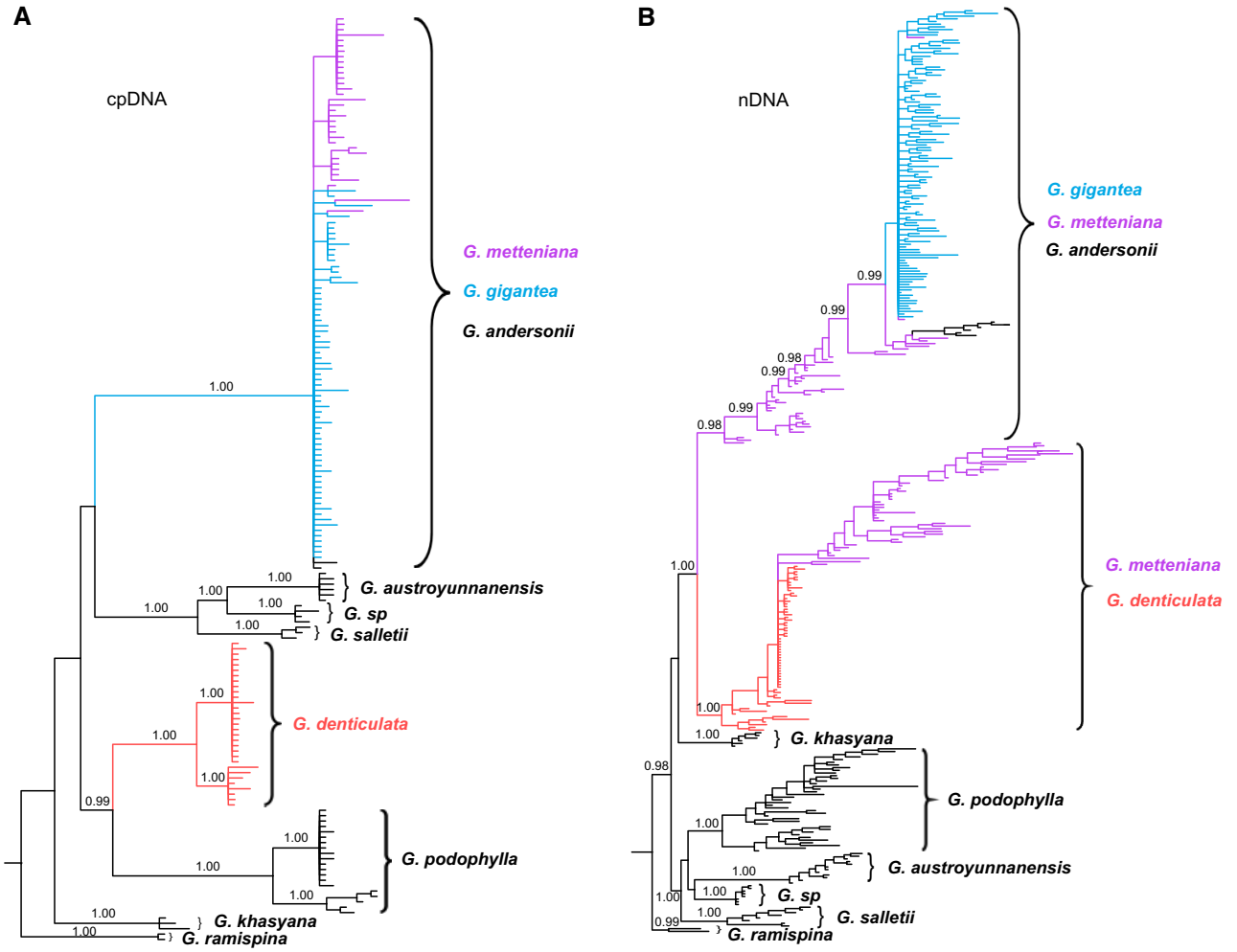

FIG. 1. Phylogenetic relationships among the focal *Gymnosphaera* species, with posterior probabilities >0.95 given at each node. (*A*) Majority rule consensus tree based on Bayesian analyses of the five concatenated cpDNA regions. (*B*) Majority rule consensus tree based Bayesian analyses of nine concatenated nDNA loci.

network (see below) appears to show that *G. gigantea* is the maternal parent of *G. metteniana*.

Sequence Variation and Haplotype Diversity

The cpDNA data set, including a total of 133 individuals from *G. metteniana* and its putative diploid progenitors (including *G. andersonii*), consisted of 72 unique haplotypes (supplementary table S4, Supplementary Material online). Of these, most were species-specific and only one haplotype (i.e., H16) was shared between *G. metteniana* and *G. gigantea* (fig. 2*A* and supplementary fig. S4, Supplementary Material online). This shared haplotype was the most common haplotype of *G. gigantea*, accounting for ca. 14% of the total individuals analyzed. The haplotype network divided the haplotypes into two separate haplo-groups: the *G. denticulata* group and the *G. metteniana*–*G. gigantea*–*G. andersonii* group (fig. 2*A*). No shared haplotype was observed between *G. metteniana* and *G. andersonii*.

Nine single-copy nuclear loci yielded an alignment that was 5,455 bp in length representing 403 individuals from 55 populations of the three species. Based on the nuclear DNA, *G. metteniana* harbors more nucleotide diversity (π) than both of the diploid progenitors at each of the nuclear loci, with estimates of π ranging from 0.0027 to 0.0109 in *G. metteniana*, 0.0007 to 0.0022 in *G. denticulata*, and 0.0011 to 0.0053 in *G. gigantea* (table 3). We consistently detected a higher level of haplotype diversity in *G. metteniana* (H_d =0.651–0.893) than both of its putative progenitors at seven of the nine nuclear loci (table 3). This is expected as the tetraploid genome of *G. metteniana* represents a combination of the diverse genomes of its parental species. Individually, most loci displayed significant negative Tajima's *D* (table 3), which might indicate population expansion in their evolutionary history. Finally, the super-network analyses of the concatenated data set of the nine nuclear loci reveals a major split between *G. denticulata* and *G. gigantea*, whereas the two homologs in *G. metteniana* grouped with each of these two clusters, respectively (fig. 2*B*).

Genetic Differentiation and Population Structure

Population genetic analyses from genomewide SNP data suggest that *G. metteniana* is a genetic intermediate compared

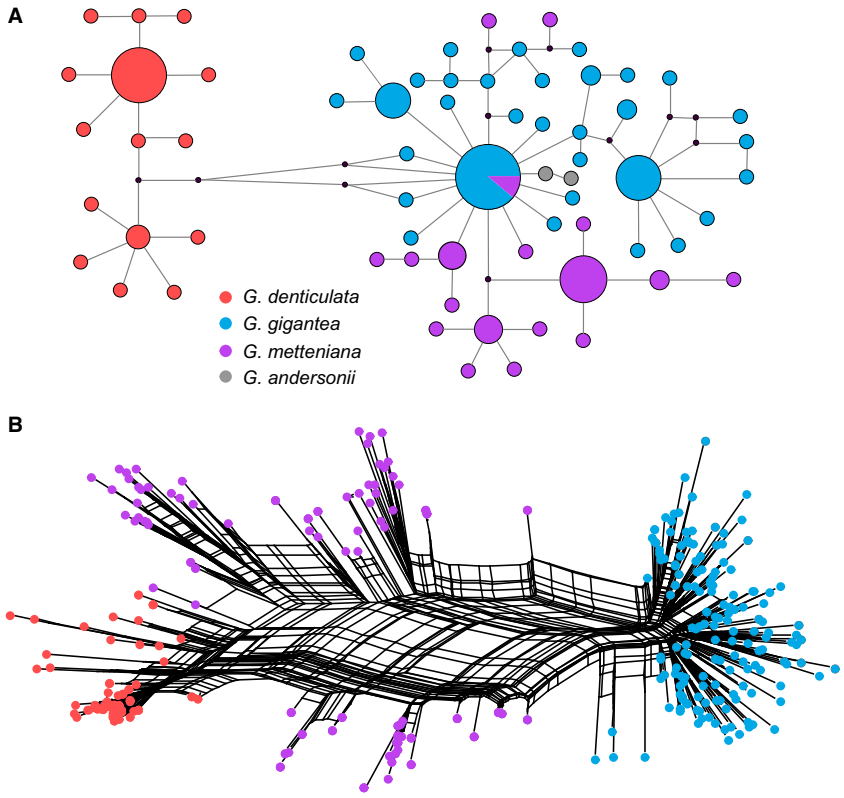

FIG. 2. (*A*) Haplotype network of the five concatenated cpDNA regions. Size of each circle indicates the approximate number of individuals in which a particular haplotype was observed. (*B*) Neighbor-net tree of 294 alleles (with two alleles for each individual) of the three species of *Gymnosphaera* based on nDNA data. Branch lengths are proportional to absolute distances calculated from the binary matrix.

with its putative parents. The genome-level *F*-statistics demonstrated high genetic differentiation between *G. denticulata* and *G. gigantea* (F_{ST}=0.834; fig. 3A). These analyses indicated that *G. metteniana* possess an intermediate gene pool relative to *G. denticulata* and *G. gigantea*, but that it is genetically slightly more similar to *G. denticulata* (*G. denticulate* × *G. metteniana* F_{ST}=0.348; *G. metteniana*×*G. gigantea* F_{ST}=0.417; fig. 3A). Consistent with this statistics, principle components analysis (PCA) based on genomic SNP recovered three well-differentiated clusters with *G. metteniana* in an intermediate position between its putative parents along the first principal coordinate (PC1; 34.8%; fig. 3B). These results together indicated that *G. metteniana* is a hybrid of *G. denticulata* and *G. gigantea* and that it has become genetically differentiated from these parental species following its origin. Similar patterns of population differentiation were observed in the nDNA data set, where mean F_{ST} between *G. denticulata* and *G. gigantea* was 0.758, which was much higher than the mean F_{ST} between *G. metteniana* and the two parents (*G. denticulate* ×*G. metteniana* F_{ST}=0.311; *G. metteniana*×*G. gigantea* F_{ST}=0.387; supplementary table S5, Supplementary Material online). In both SNP and nDNA sequence data sets, population differentiation within species was fairly low in *G. metteniana* (F_{ST}=0.049 for SNP and 0.076 for nDNA) compared with *G. denticulata* (F_{ST}=0.319 for SNP and 0.244 for nDNA) and *G. gigantea* (F_{ST}=0.208 for SNP and 0.147 for nDNA) (fig. 3A and supplementary table S5, Supplementary Material online).

To assess the overall genomic composition of the *G. metteniana* populations, we performed a Bayesian analysis in STRUCTURE of the genomewide SNPs to cluster *G. denticulata*, *G. gigantea*, and *G. metteniana* individuals under the assumption that they represented admixed populations. Under the best-fit model of $K = 2$ (fig. 3C), the genomes of all individuals from *G. denticulata* and *G. gigantea* clustered to different groups with high probability (100%), whereas *G. metteniana* shares an average of ca. 60% of its genome with *G. denticulata* and 40% with *G. gigantea* (fig. 3C). Where $K = 3$, *G. metteniana* formed its own cluster, although with substantial admixture from both of its parents. Overall, these results from population genetic analysis revealed that *G. metteniana* had widespread genomic admixture indicative of hybrid origin. STRUCTURE analysis on the nDNA data set revealed a similar pattern of genetic composition of the *G. metteniana* populations (fig. S5, Supplementary Material online). At $K = 2$, we observed a pattern of varied *G. denticulata* ancestry in the *G. metteniana* populations according to spatial distribution. Specifically, the genomes of *G. metteniana* individuals from parapatric/sympatric populations with *G. denticulata* show higher proportions of *G. denticulata* ancestry (72% for SNP and 70% for nDNA) than those from allopatric populations (57% for SNP and

Table 3. Nucleotide Diversity Levels and Tajima's *D* of the Combined cpDNA Sequence and Nine nDNA of the Studied Species.

Locus	Species	*N*	*S*	*H*	H_d	π	Tajima's *D*
cpDNA	*Gymnosphaera denticulata*	30	2	15	0.784	0.0017	1.459
	Gymnosphaera metteniana	39	38	21	0.935	0.0022	−2.291**
	Gymnosphaera gigantea	62	19	34	0.912	0.0006	−2.077*
5229	*Gymnosphaera denticulata*	77	13	13	0.326	0.0011	−1.832*
	Gymnosphaera metteniana	105	15	30	0.813	0.0075	1.510
	Gymnosphaera gigantea	203	33	31	0.394	0.0012	−2.379**
5320	*Gymnosphaera denticulata*	73	11	11	0.322	0.0018	−0.927
	Gymnosphaera metteniana	111	18	35	0.797	0.0073	1.737
	Gymnosphaera gigantea	197	63	100	0.932	0.0053	−1.911*
5521	*Gymnosphaera denticulata*	78	14	13	0.576	0.0021	−1.110
	Gymnosphaera metteniana	111	20	34	0.893	0.0055	0.253
	Gymnosphaera gigantea	206	45	68	0.828	0.0042	−1.759*
5604	*Gymnosphaera denticulata*	78	10	10	0.340	0.0007	−1.761
	Gymnosphaera metteniana	112	8	10	0.651	0.0031	1.292
	Gymnosphaera gigantea	208	29	30	0.486	0.0011	−2.191**
5770	*Gymnosphaera denticulata*	75	12	10	0.359	0.0015	−1.281
	Gymnosphaera metteniana	114	11	18	0.744	0.0028	0.142
	Gymnosphaera gigantea	195	24	39	0.760	0.0023	−1.503
5839	*Gymnosphaera denticulata*	79	7	5	0.258	0.0022	−0.249
	Gymnosphaera metteniana	109	25	30	0.786	0.0109	0.804
	Gymnosphaera gigantea	191	39	36	0.566	0.0021	2.301**
5894	*Gymnosphaera denticulata*	80	18	15	0.493	0.0016	−1.907*
	Gymnosphaera metteniana	113	12	19	0.760	0.0027	−0.333
	Gymnosphaera gigantea	207	28	39	0.692	0.0018	−1.995*
6148	*Gymnosphaera denticulata*	78	2	4	0.244	0.0011	−0.435
	Gymnosphaera metteniana	98	13	23	0.822	0.0074	−0.114
	Gymnosphaera gigantea	199	15	16	0.429	0.0018	−1.825*
6318	*Gymnosphaera denticulata*	74	13	15	0.401	0.0014	−1.622
	Gymnosphaera metteniana	107	19	30	0.827	0.0054	−0.172
	Gymnosphaera gigantea	184	25	46	0.884	0.0032	−1.336

NOTE.—*N*, number of individuals; *S*, number of segregating sites; *h*, number of haplotypes; H_d, haplotype diversity; π, nucleotide diversity.
*$0.01 < P < 0.05$; **$0.001 < P < 0.01$.

58% for nDNA; fig. 3*D*), with the *P* value (Wilcoxon tests) being 0.001 and 0.025 for SNP and nDNA, respectively.

Modeling Postpolyploidization Gene Flow and Population Growth

The allopolyploid formation for *G. metteniana* has either resulted from a single hybridization event or multiple events, and its history may include postspeciation gene flow or not. To test these alternatives, we used coalescent modeling in a composite likelihood framework to compare the fit of 23 speciation models (fig. S6, Supplementary Material online) for the origin of *G. metteniana*. This analysis showed that a scenario involving a single hybridization event (H1) between *G. denticulata* and *G. gigantea* as the most likely hypothesis explaining the origin of *G. metteniana* and was superior to a hypothesis with multiple hybridization events (H2) ($\Delta i > 15{,}000$; supplementary table S6, Supplementary Material online). Further analysis showed that the model (A3) assuming both historical and current gene flow occurring between *G. metteniana* and both parent species performed better than between *G. metteniana* and *G. denticulata* or *G. gigantean* only (supplementary table S6, Supplementary Material online). However, it has to be noted that models assuming recent gene flow (A2) or constant gene flow (A3) between *G. metteniana* and both parent species performed similarly ($\Delta i < 10$; supplementary table S6, Supplementary Material online). Hence, we further tested for changes in population size based on the two best migration models. Further analysis demonstrated that the model assuming an exponential growth since the origin of the three species together with constant gene flow (G2) clearly performed better than models assuming a recent population exponential growth or a sudden change in population size (either a bottleneck or a sudden expansion).

The estimated parameters of the best model overall suggest that *G. metteniana* originated around 0.83 to 4.61 Ma (fig. 4), whereas the divergence between *G. denticulata* and *G. gigantea* was estimated to have occurred in 2.15–9.83 Ma. Under this scenario, the initial parental contributions in the nuclear genomic composition of *G. metteniana* population were 53.04% for *G. denticulata* and 46.96% *G. gigantea* (fig. 4), respectively. The better fit of a hybrid speciation model with asymmetric gene flow indicated that ongoing gene flow may occur, with higher migration rates between *G. metteniana* and *G. denticulata* (*Nm*=0.161–0.174) than between *G. metteniana* and *G. gigantea* (*Nm*=0.085–0.133). Additionally, the best model suggested negative exponential growth in the three *Gymnosphaera* species despite that the growth rates were relatively small (2.80×10^{-8}, 3.14×10^{-8}, and 1.50×10^{-6} for *G. denticulata*, *G. metteniana*, and *G. gigantea*, respectively).

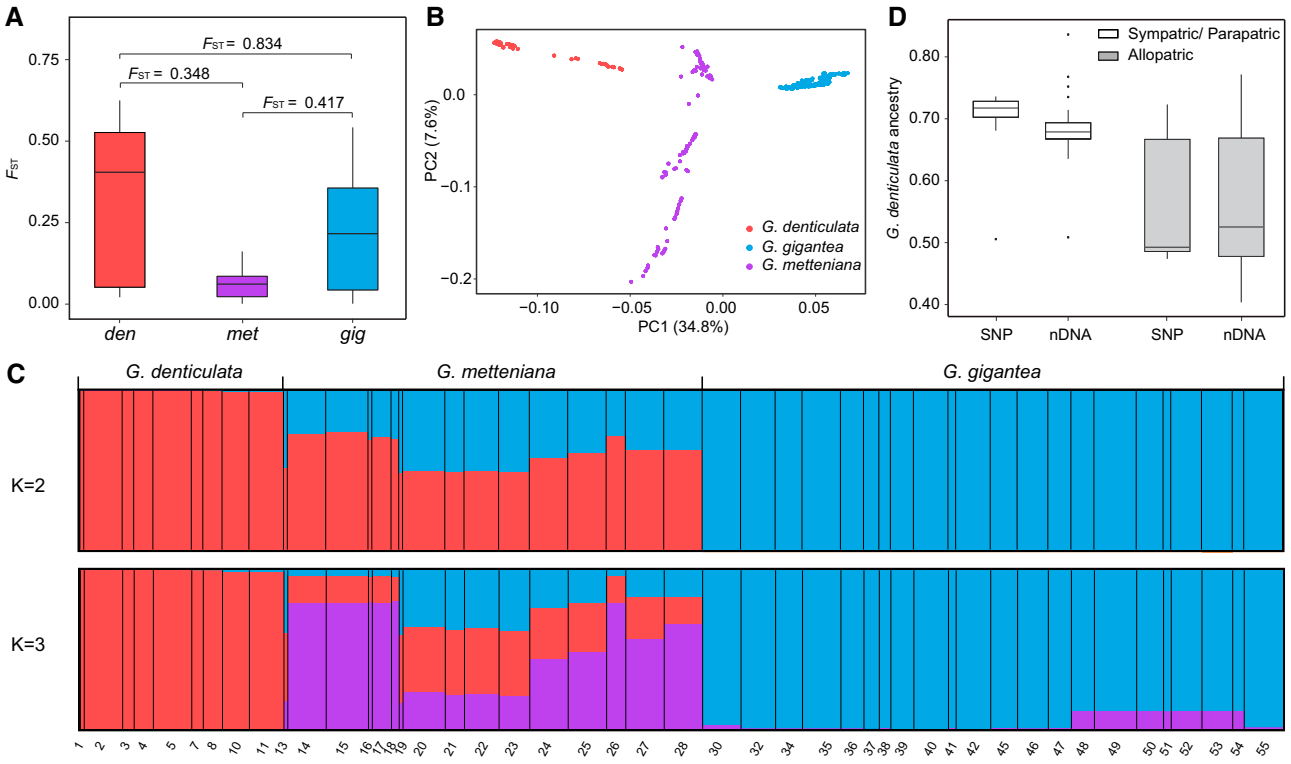

Fig. 3. Estimated population structure in the three species of *Gymnosphaera*. (*A*) Genetic differentiation (F_{ST}) within each species and between each of the three species pairs based on genomewide SNP data (species abbreviation: *dent* for *Gymnosphaera denticulata*, *met* for *G. metteniana*, and *gig* for *G. gigantea*). (*B*) Plot of the first two dimensions of PCA for the three species of *Gymnosphaera*, with the first two axes (PC1 and PC2) explaining 34.8% and 7.6% of the variation, respectively. (*C*) The results for STRUCTURE analysis with $K = 2$ and $K = 3$ based on genomewide SNP data. Populations are ordered according to the localities listed in supplementary table S9, Supplementary Material online. (*D*) Box plots for the proportion of *G. denticulata* ancestry in its sympatric/parapatric and allopatric *G. metteniana* populations based on SNP and nDNA data sets.

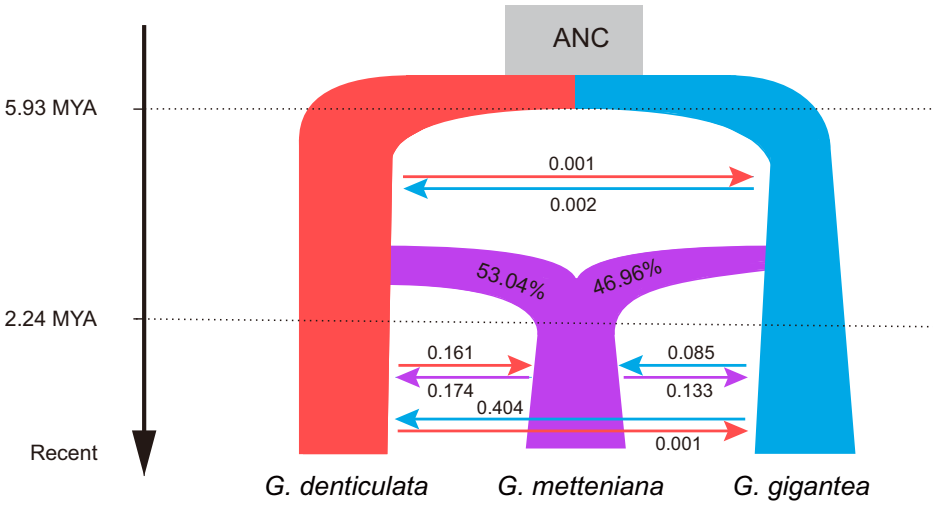

Fig. 4. Maximum-likelihood parameter estimates of the best-supported demographic history model (G2) of *Gymnosphaera metteniana*, *G. denticulate*, and *G. gigantea*. Arrows indicate migration of individuals (*Nm*) per generation between the three species. The percentage indicates nuclear genomic compositions from parent species to the hybrid species. Line width represents effective population sizes. The left axis shows the timescale in units of millions of years before present. ANC is the effective population size of the common ancestor of the three species.

Evidence of Ecological Niche Divergence

Ecological niches inferred in MAXENT showed good ability to predict the distributions of each species, with the values of mean area under the receiver operator curve being 0.948 ± 0.020, 0.932 ± 0.034, and 0.892 ± 0.054 for *G. denticulata*, *G. metteniana*, and *G. gigantea*, respectively. Mean regularized training gain across the ten replicates was 2.148, 1.992, and 1.836, with the corresponding suitability threshold containing 90% of training samples being 0.236, 0.291, and 0.223 for *G. denticulata*, *G. metteniana*, and

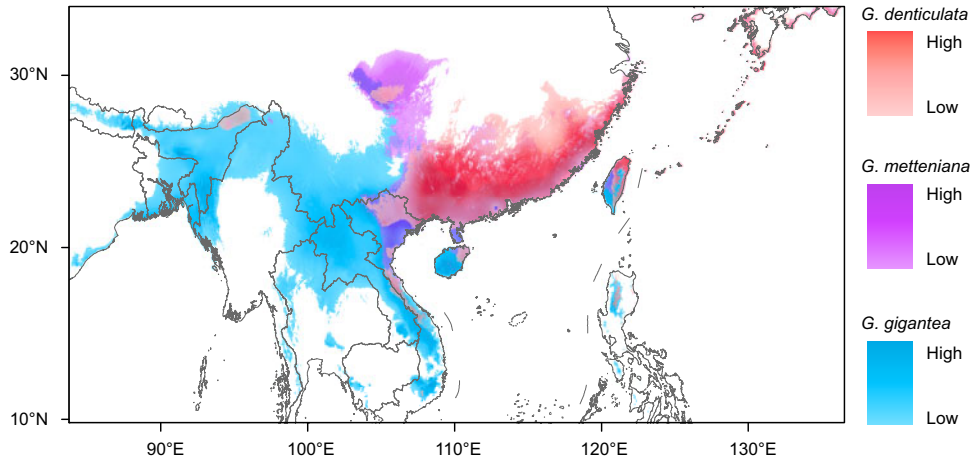

FIG. 5. Predicted distributions of the three species of *Gymnosphaera* based on ecological niche modeling using Maxent.

G. gigantea, respectively. Predicted distributions were generally congruent with the observed distribution of each species (fig. 5 and supplementary fig. S7, Supplementary Material online). Precipitation of the driest quarter contributed most to the model predictions for both *G. denticulata* (60.3%) and *G. metteniana* (43.8%), whereas temperature annual range (i.e., difference between maximum temperature of the warmest month and minimum temperature of the coldest month) contributed most to the model predictions of *G. gigantea* (57.0%) (supplementary table S7, Supplementary Material online). Niche overlap statistics demonstrated that *G. gigantea* occupied a distinct niche, whereas the niches between *G. metteniana* and *G. denticulata* were not significantly more different than expected by chance (supplementary table S8, Supplementary Material online). Niche equivalency tests revealed statistically significant niche divergence for all comparisons ($P < 0.05$), except the statistic for *G. denticulata* vs. *G. metteniana* ($P = 0.06$) (fig. S8, Supplementary Material online).

Discussion

Allopolyploid Origin of *G. metteniana*

Whereas polyploids are well documented in ferns, and the probability of hybridization as a speciation mechanism has long been recognized (Sigel 2016), our study provides the first genomewide assessment of hybrid speciation and introgression with statistical model tests for this vascular plant lineage. We show that tetraploid *G. metteniana* originated via hybridization between *G. gigantea* and *G. denticulata*, and we also show that *G. gigantea* might have served as the maternal parent and *G. dentitulata* as the paternal parent. Overall, we present a first well-documented case of polyploid speciation in the family Cyatheaceae, although polyploid speciation has been reported in many other fern lineages based on phylogenetic and/or morphological evidences (e.g., Perrie and Brownsey 2005; Sessa et al. 2012; Rothfels et al. 2014; Schneider et al. 2017; Dauphin et al. 2018).

Although polyploids have been rarely reported in the family Cyatheaceae (Nakato 1989), natural hybridization events are presumably frequent and have been suggested in numerous instances for species occurring in the subtropics and tropics of the world (Conant 1975; Conant and Cooper-Driver 1980; Edwards 2005; Janssen and Rakotondrainibe 2007). Specifically, Conant and Cooper-Driver (1980) proposed that autogamous allohomoploid speciation may explain stable hybrid populations of crosses among three sympatric species of *Alsophila*. Under this hypothesized mode of speciation, homoploid hybrid species can arise when allohomoploid F1 hybrids are fertile selfers that quickly give rise to a colony of homozygous genetically and morphologically identical sporophytes. However, this hypothesis of homoploid hybrid speciation can be easily dismissed in the case of *G. metteniana* because it is a tetraploid as demonstrated in our study and that of Nakato (1989). Nevertheless, allohomoploid hybrids could have been intermediate to the formation of allopolyploids (Sigel 2016). In this scenario, an allohomoploid hybrid could have produced unreduced fertile spores, which underwent subsequent intragametophyic selfing to produce a hybrid species with chromosome doubling. In an alternative scenario, allopolyploids arose through the direct fusion of unreduced gametes of each species; that is, a triploid bridge (Ramsey and Schemske 1998; Sigel 2016). In flowering plants, this triploid bridge pathway has been proposed as the most extended route to allopolyploid speciation (Ramsey and Schemske 1998; Husband 2004). Although we cannot determine which mechanism of allopolyploid formation is responsible for the origin of *G. metteniana* with our available data, the prevailing production of unreduced diploid gametes in homosporous ferns (Haufler 2002) suggests that *G. metteniana* is likely to have arisen from intragametophyic selfing of an intermediate homoploid hybrid or via direct union of two unreduced gametes. These pathways to allopolyploids have also been proposed in the formation of other polyploid ferns (Kawakami et al. 2010; Hunt et al. 2011).

The comparison of phylogenies from the cpDNA and nDNA data sets suggests that hybridization between the two progenitors of *G. metteniana* occurred unidirectionally. Such patterns of nonreciprocal hybridization have been reported to occur in several fern genera such as *Asplenium*

(Perrie et al. 2010; Hunt et al, 2011), *Acrostichum* (Zhang et al. 2013), and *Dryopteris* (Sessa et al. 2012; Testo et al. 2015). Recent simulation studies demonstrated that fixation of genetically incompatible loci from each parental species can readily lead to the evolution of reproductive isolation in hybrid populations and serve as a crucial step toward hybrid speciation (Schumer et al. 2015; Comeault 2018). The pattern of unidirectional hybridization inferred in *Gymnosphaera* may be caused by nuclear-cytoplasmic incompatibilities (Nakazato et al. 2007). If this were the case, it might be that evolution of reproductive isolation in *G. metteniana* was rapid as a result of mosaic inheritance of different parental alleles at incompatible loci, enforcing reproductive barriers.

Many polyploid species of flowering plants are known to arise from multiple hybridization events (Soltis et al. 2014), and this "multiple origin" phenomenon has also been reported in ferns (Perrie et al. 2010; Hunt et al. 2011; Beck et al. 2012; Sigel et al. 2014; Fujiwara et al. 2018). Multiple independent origins leading to multiple lineages may be possible in *G. metteniana* especially based on its disjunct geographical distribution. If so, there would be evidence of divergent cpDNA sequence and nuclear genome among populations of the hybrid species, in parallel to the observations in its parent species. In contrast, patterns of variation in cpDNA and genomewide SNPs show that different populations of *G. metteniana* are quite genetically homogeneous, with intraspecific genetic differentiation (F_{ST}) in the hybrid species that is relatively low compared with its parents (fig. 3*A*). In addition, only one shared chloroplast haplotype was identified between *G. metteniana* and *G. gigantea* (fig. 2*A*). This single shared haplotype is mainly detected in *G. gigantea* populations from southern China. Our modeling analysis further demonstrated a single origin of *G. metteniana*. Nevertheless, our results could not efficiently discern a scenario in which the origin of *G. metteniana* may have taken place via a single hybridization event from a scenario involving multiple polyploidization events in a highly narrow region from South China, especially when their distribution areas still overlap. The presence of many private haplotypes with low frequency in the hybrid species indicates that they have arisen by mutation from the shared haplotypes after polyploidization, whereas the current wide geographical distribution likely resulted from range expansion following the establishment of the hybrid species (see below).

Allopolyploid Speciation Accompanied by Gene Flow

Polyploid speciation is considered to be a special form of speciation in which whole-genome changes result in rapid, strong interploid reproductive isolation (Coyne and Orr 2004). The STRUCTURE analyses revealed that *G. metteniana* individuals that are sympatric/parapatric with *G. denticulata* have significantly higher shared ancestry than individuals from allopatric populations (fig. 3*D*). Thus, it is probable that there has been recent introgression into the tetraploid, *G. metteniana*, from the diploid, *G. denticulata*. Coalescent modeling further supported a scenario of ongoing gene flow between the hybrid species and its progenitors (fig. 4). Such a pattern of interploid gene flow can occur through either triploid bridges or unreduced gametes (Petit et al. 1999). In our data, the gene flow into the tetraploid *G. metteniana* is more likely through unreduced gametes of the diploid *G. denticulata* rather than via triploid bridges, because of the lack of triploid hybrids detected in this study (table 2). In addition to the role in the formation of many polyploid lineages (Ramsey and Schemske 2002), unreduced gametes have also been considered to facilitate interploid gene flow within and between species (Sora et al. 2016). In flowering plants, the rate of unreduced gamete formation was estimated to vary from 0.0% to 81.1% of all gametes produced (Ramsey and Schemske 1998). By comparison with flowering plants, ferns may experience a higher rate of unreduced gametes due to their greater number of meiotic events (Haufler 2002). Although rates of unreduced gamete formation have not been studied in *Gymnosphaera* species, we suggest that unreduced gametes may have played an important role in mediating reproduction between the polyploid species and its progenitors. Evidence of interploid gene flow has been demonstrated in many angiosperms (Menken et al. 1995; Slotte et al. 2008; Ma et al. 2010; Jørgensen et al. 2011; Han et al. 2015; Sutherland and Galloway 2017). To our knowledge, only one other study has analyzed progenitor-to-polyploid introgression in ferns, but this study used amplified fragment length polymorphism markers and revealed little evidence of introgression from progenitors to polyploids of *Asplenium* species (Perrie et al. 2010).

Despite the evidence for high levels of interploid gene flow, the tetraploid *G. metteniana* comprises a genomically distinct entity from its progenitors. This is in line with a mode of speciation with gene flow, which predicates that speciation is plausible if divergent natural selection is strong enough to counteract gene flow (Coyne and Orr 2004). Selection for ecological specialization between polyploid species and their diploid progenitors might limit interploid gene flow. Simulation and empirical studies demonstrated that the establishment of polyploid lineages is commonly associated with ecological divergence (Parisod 2012; Fowler and Levin 2016). If new polyploids possess novel physiological, ecological, or phenological characteristics, environmental heterogeneity could facilitate their establishment within a new niche, and create reproductive isolation from its parental species. In agreement with this predication, our MAXENT analysis identified a significant difference in the climatic niches occupied by *G. metteniana* and its parental progenitor *G. gigantea*. We found that the climatic niches of *G. metteniana* and *G. denticulata* largely overlap, although sympatry of the two species is rare in wild. This suggests that, broadly, climatic conditions were probably not determinate factors of species isolation. Instead, the observed isolation was caused either by intrinsic reproductive isolation between the hybrid species and its parental progenitors or ecological isolation over short distances caused by microhabitat preferences. In contest to the general assumptions about fern dispersal, scaly tree ferns have a low dispersal potential (Conant 1978; Bernabé et al. 1999; Ramírez-Barahona and Eguiarte 2015), which may facilitate divergent selection at local scales. Significant fine-scale spatial genetic structure (SGS) has been observed for several

Alsophila species (Conant and Cooper-Driver 1980; Ramírez-Barahona and Eguiarte 2015), and such a pattern of SGS was attributed to limited spore dispersal caused by micro-habitat spatial heterogeneity. A contributing factor is arguably the relative short period of viability of tree fern spores (Li et al. 2010). These observations are indicative of local adaptation and support the untested hypothesis that relates high local endemism in the Cyatheaceae to ecological specialization (Tryon and Gastony 1975).

Demographical History of *G. metteniana*

Consistent with the divergence time of 2.4 Ma (Dong et al. 2019) estimated from multilocus DNA sequences, the timing of the hybridization event that gave rise to *G. metteniana* is inferred to be around 2.24 Ma (95% high posterior density: 0.83–4.61 Ma), which appears to coincide with intensification of the East Asian monsoon in the Pliocene and Pleistocene periods (~3 Ma) (Wan et al. 2007). The intensified monsoon regime, which is characterized by an increased seasonality of precipitation, might have supported a high rate of speciation in southern China, including evolutionary radiation of angiosperms (Kong et al. 2017) and ferns (Wang et al. 2012). It is also likely that the intensified monsoon regime around the Pliocene and Pleistocene periods facilitated speciation in ferns via wind- and water-mediated dispersals of their spores. Furthermore, the intensified monsoon regimes may have created novel habitats for and promoted the establishment of *G. metteniana* populations. The current patterns of disjunct distributions may have resulted from either long-distance dispersal into available habitats or shift in geographical distribution during the Quaternary climate oscillations. The negative values for Tajima's *D* retrieved for most loci in the three species indicate that they are in the process of population expansion. Evidence of range expansion from our analyses of nuclear genetic data is also supported by the negative exponential growth detected in our model predictions.

Conclusions

Our data provide strong support for the allopolyploid origin of *G. metteniana* accompanied by ongoing gene flow. Particularly, we detected substantial introgression of *G. denticulate* into *G. metteniana* probably via unreduced gametes. The timing of hybridization and polyploidization likely coincided with the period of intensified East Asian monsoon around the Pliocene and Pleistocene epochs in southern China. Although we demonstrated ecological niche separation between the allotetraploid *G. metteniana* from its diploid progenitor, *G. gigantea*, the similar niche observed between *G. metteniana* and *G. denticulata* suggests that either some degree of intrinsic genetic isolation or local adaptation may be playing an important role in the evolution of reproductive barriers. To our knowledge, our study represents one of the best-documented cases thus far of allopolyploid speciation in ferns. Further studies are needed to investigate the genetic and ecological mechanisms underlying reproductive isolation between the hybrid species and its parent species. Such information could be useful for understanding the relative contributions of gene flow and divergent selection in the stabilization of hybrid polyploid species. Overall, our empirical study strengthens the argument that speciation with gene flow is frequent and can even be found in the process of allopolyploid speciation.

Materials and Methods

Flow Cytometry Analysis

We performed flow cytometry to compare genome sizes in *Gymnosphaera metteniana*, *G. denticulata*, and *G. gigantea* and to estimate ploidy levels based on samples from 11 natural populations of the three species (three, three, and five populations, respectively) plus three living collections from the South China Botanical Garden (SCBG) (supplementary table S1, Supplementary Material online). The number of plants analyzed per population varied from one to four individuals. Young, fresh leaves were obtained from individuals within the populations. These were stored at 4°C during field collections and immediately returned to the lab thereafter for experimental analysis. Because our preliminary survey revealed that 1C-value of the three species ranged between 7.0 and 15.0 pg, we selected *Nicotiana tabacum* (1C = 4.5 pg, Leitch et al. 2008) as an appropriate reference standard. We extracted the stained sampled using Sartec CyStain PI Absolute P kit (Partec GmbH, Münster, Germany) and then analyzed the samples of a Partec CyFlow Space cytometer (Partec GmbH), on which we recorded the fluorescence intensity of 8,000–10,000 particles. The genome sizes are shown as the complete 2C-DNA content of the nucleus, irrespective of ploidy. For each sample, we estimated the genome size using *G. gigantea* as a standard. Additionally, we sought to directly determine the ploidy of each sample following a similar protocol as mentioned above, except that we used solution A of the High Resolution Kit (Partec GmbH) to obtain the stained cells. Again, we used *G. gigantea* as an internal standard.

Sample Collection, DNA Extraction, and Sequencing

We collected 541 samples from 92 populations (1–12 individuals per population) of the 10 *Gymnosphaera* species, three *Alsophila* species and *Sphaeropteris brunoniana* (supplementary table S9, Supplementary Material online). Of these, 55 populations belonged to our ingroup: 12 *G. denticulata*, 16 *G. metteniana*, and 27 *G. gigantea* (supplementary fig. S1 and table S9, Supplementary Material online). We aimed to include populations representing the entirety of the distribution of the ingroup with the exception of populations of *G. gigantea* in Laos and Thailand. We initially identified samples based on their morphology, and voucher specimens are deposited in the Herbarium of South China Botanical Garden (IBSC). We extracted total genomic DNA from the silica-dried leaves using a modified cetyltrimethylammonium bromide method (Doyle JJ and Doyle JL 1987).

To determine which species served as parents in the origin of *G. metteniana*, we reconstructed phylogenies of the genus *Gymnosphaera* using cpDNA and nDNA sequence data. We chose a subset of 175 individuals from 92 populations (one to three individuals per population) (supplementary table S9,

Supplementary Material online) and sequenced five cpDNA regions (*rbcL*, *rbcL-accD*, *rbcL-atpB*, *trnG-trnR*, and *trnL-trnF*; supplementary table S2, Supplementary Material online) and nine nDNA loci (supplementary table S3, Supplementary Material online). For population genetic analysis, we sequenced the nine nDNA loci in a total of 403 individuals from 55 populations of the three species (*G. denticulata*, *G. metteniana*, and *G. gigantea*). We derived protocols for amplification of the cpDNA markers via polymerase chain reaction (PCR) from Dong and Zuo (2018), and primers are listed in supplementary table S3, Supplementary Material online. We followed Dong et al. (2019) in determining nuclear markers known to be variable in Cyatheaceae and performing suitable PCR amplification protocols. We deposited all sequences generated in this study in GenBank (supplementary tables S10 and S11, Supplementary Material online).

We then chose a second subset of 331 individuals for restriction-site-associated DNA sequencing (RAD-seq; Baird et al. 2008). RAD libraries were constructed and sequenced by Novogene Bioinformatics Institute (Beijing, China) and run on an Illumina HiSeq 2000 platform (San Diego, CA) with 150-bp single-end reads. After excluding individuals with low-quality sequence, the sample set used for population genomic analysis contained 313 individuals of the three species, including 54 *G. denticulata*, 108 *G. metteniana*, and 151 *G. gigantea* (supplementary table S9, Supplementary Material online).

Phylogenetic Analysis with cpDNA and nDNA Sequence Data

We reconstructed independent phylogenetic trees for concatenated data sets of cpDNA and nDNA. We performed phylogenetic reconstruction using Bayesian inference (BI) implemented in MrBayes 3.2.6 (Ronquist et al. 2012) on the high performance computing infrastructure via the CIPRES Science Gateway 3.3 (Miller et al. 2015). The concatenated length for five cpDNA fragments was 4,157 bp, and the matrix contained 612 variable sites in total. For the nDNA sequences, we phased alleles statistically in PHASE 2.1.1 (Stephens et al. 2001) using input files assembled on the SeqPHASE web server (Flot 2010). The concatenated alignment after phasing was 5,470 bp long, of which 570 cites were parsimony-informative (supplementary table S2, Supplementary Material online). We determined the best-fit model of evolution for each locus using the Akaike information criterion implemented in MrModeltest 3.7 (Posada and Crandall 1998), which is incorporated in PAUP* 4.0b10 (Swofford 2002). We performed two independent BI runs with one cold and three heated chains for 80 million Markov chain Monte Carlo generations. We sampled trees every 8,000 generations and discarded the first 25% generations as burn-in. We checked for convergence between runs and stationarity based on effective sample size in Tracer 1.7 (Rambaut et al. 2018). We calculated the support values with 1,000 bootstrap replicates, each with 10 random sequence addition replicates holding one tree per replicate, Tree Bisection-Reconnection branch swapping, and Multrees on. In all phylogenetic analyses, we used *Sphaeropteris brunoniana* as the outgroup (Korall et al. 2007; Dong and Zuo, 2018).

Sequence Diversity and Haplotype Network Analysis

Because chloroplast fragments are effectively inherited as one locus in plants, they were combined into a single locus for subsequent analysis. We determined haplotypes (h), segregating sites (S), haplotypic diversity (H_d), nucleotide diversity (π, Nei 1987), and Tajima's D to detect signatures of past demographic events (Tajima 1989) separately for the three species at each nuclear locus and the cpDNA in ARLEQUIN 3.5 (Excoffier and Lischer 2010). Pairwise F_{ST} (Wright 1984) was used to measure population differentiation within and between species as implemented in ARLEQUIN and the significance of the observed F_{ST} was tested using 10,000 random permutations of the data matrix. A haplotype network based on cpDNA data was reconstructed using a maximum parsimony method based on a median-joining algorithm as implemented in the software NETWORK 4.6 (Bandelt et al. 1999). To visualize the pattern of genetic clustering in the nDNA data, we produced a NeighbourNet diagram based on the uncorrected P-distance between individuals with SplitsTree4 (Huson and Bryant 2006).

Bioinformatics Treatment and SNP Calling

It is a challenge to screen SNPs for polyploid species. However, allopolyploids usually undergo a series of homolog silencing and loss over time that ultimately lead to genomic diploidization (Doyle et al. 2008). Given that it has been a long time of >2 My since the formation of *G. metteniana*, we assumed an effectively diploid genetic system for this species and called SNPs as we did for its diploid parents. Briefly, raw sequence reads were processed using STACKS 1.47 software pipeline (Catchen et al. 2013) to obtain SNPs and genotype data sets. Reads were firstly filtered for quality by identifying and removing PCR duplicates, and discarding reads with no correct barcode and the *Eco*RI recognition site using STACKS. Nucleotide base calls with a Phred quality score below 20 were replaced with "N"s, and only reads with <5% Ns were retained. After filtering, 9,182,555,777 clean reads were retained across 313 individuals. The number of sequence reads per sample ranged between 21,188,150 and 41,507,479, with the median value being 29,229,964 (supplementary table S12, Supplementary Material online). The retained clean reads were further processed in ustacks. We set the minimum depth of coverage to create a stack at five sequences and an alpha value of 0.05 for the SNP model. We constructed a catalog of consensus loci containing all the stacks from all the individuals and merged all alleles together with cstacks. Then we compared each individual genotype against the merged catalog using sstacks. Finally, we used the populations program to obtain the loci that were present in at least 80% of the individuals from each population in at least 24 populations with at least five RAD tags per allele at each locus (5× coverage per allele). We kept only the first SNP per locus and removed loci with minor allele frequencies <0.02. We removed loci with extremely high coverage (coverage greater than 2SD above the mean) and discarded samples with low genotype coverage rate (<60%) using VCFTOOLS (Danecek et al. 2011). Using this pipeline, we obtained Data set 1 with a total of 13,739 biallelic SNPs for population

genetic analyses. Mean coverage per locus across individuals ranged from 4.3 to 50.8 (median = 10.1; supplementary fig. S9, Supplementary Material online), and the number of loci per individual varied between 4,624 and 11,792 (median = 8,518, i.e., about 62% all available SNPs; supplementary fig. S10, Supplementary Material online).

Additionally, we constructed two more data sets with different SNP calling and filtering strategies aiming for other analyses. Specifically, Data set 2 was filtered for calculating pairwise genetic differentiation among populations. To avoid the potential impact of different sample sizes, we only kept populations with at least five samples and randomly selected five samples per population. We reran the populations program as above and the resulting Data set 2 has 6,062 SNPs with locus coverage of at least 80% for 179 individuals. Based on Data set 1, we further filtered SNPs for coalescent modeling by selecting one individual with the least amount of missing data from 40 populations (9 *G. denticulata*, 15 *G. gigantea*, and 16 *G. metteniana*), pooling them by species and only retaining SNPs with no missing data. The resulting Data set 3 had 324 loci for 40 individuals in total.

Population Structure and Admixture

We calculated genomewide pairwise Hudson's F_{ST} at the species level using custom scripts from Barrera-Guzmán et al. (2018) with Data set 1. The 95% confidence intervals were obtained with 1,000 bootstraps. Pairwise F_{ST} among populations were analyzed with Data set 2. We further examined population genetic structure using the Bayesian clustering approach implemented in STRUCTURE v2.3.4. (Pritchard et al. 2000). Ten independent runs with 1,000,000 replicates after a 500,000 burn-in were conducted for each *K*, which was tested from 1 to 5. The admixture model and correlated allele frequencies between populations were specified for each run. We identified the most likely *K* using the delta *K* method of Evanno et al. (2005) with software STRUCTURE HARVESTER (Earl and Vonholdt 2012). The proportion of membership to each of the clusters was obtained by combining results across replicate runs using the program CLUMPP (Jakobsson and Rosenberg 2007). We visualized the final output of structure analysis with DISTRUCT v1.1 (Rosenberg 2004). A PCA was employed using the SNPRelate package (Zheng et al. 2012) in R Core Development Team (2010) to further examine the population structure. Both STRUCTURE and PCA analyses were performed with Data set 1.

Modeling Gene Flow and Demographical History

We used coalescent modeling in a composite likelihood framework implemented in Fastsimcoal v2.6 (Excoffier et al. 2013) to compare different evolutionary scenarios for the origin of *G. metteniana* using Data set 3. A total of 23 scenarios (supplementary fig. S6, Supplementary Material online) were designed to test different hypotheses for the origin of *G. metteniana*, the presence/absence of gene flow and its tempo and directionality, and the changes in effective population sizes during divergence. The main considerations of these models include: (1) Does *G. metteniana* originate from a single hybridization event or multiple independent events? (2) Does gene flow occur between the tetraploid *G. metteniana* and its putative diploid progenitors? (3) Does postpolyploidization gene flow occur only historically, or only recent gene flow, or both? Model H1 represented a hybrid speciation model without gene flow in which *G. metteniana* originates following admixture between *G. denticulata* and *G. gigantea*, whereas model H2 involved two independent hybridization events between *G. denticulata* and *G. gigantea* giving rise to *G. metteniana*. We also tested nine models (A1–A9) with asymmetric migration to examine if constant gene flow occurs, or gene flow is limited to a specific postpolyploidization time period in the past, or only occurs recently. For postpolyploidization gene flow, we also tested if it occurred between *G. metteniana* and both ancestral species, between *G. metteniana* and *G. denticulata* only, and between *G. metteniana* and *G. gigantea* only. Finally, we tested for changes in effective population sizes using several models with asymmetric migration events. Models G1–G4 represented population exponential growth based on the best two migration models (A2–A3), respectively, whereas models E1–E4 and B1–B4 represented a sudden population expansion and bottleneck, respectively. Among them, models G1–G2, E1–E2, and B1–B2 were used to test the continuous changes in population size as the origins of the three species, whereas the others represented recent changes in population size.

We obtained the folded site frequency spectrum data with ARLEQUIN (Excoffier and Lischer 2010). We assumed a mutation rate of 6.8×10^{-10} following a genomewide comparative study performed in ferns (Grusz et al. 2016) and a generation time of 5 years to calibrate our model. A total of 27 parameters were used to compare the observed and simulated data comparisons. For each scenario, 50 independent runs were carried out with the following settings: -n 100,000 -m -M -L 40 -q. Each run started from a different set of random starting parameters drawn from uniform and log-uniform distributions, and the data were modeled as FREQ. For each prior, a log-uniform distribution with large interval (10^2 to 10^6) was set for effective population sizes due to lack of knowledge concerning population sizes. We set all the time intervals for the species divergence according to our molecular dating analysis in the taxa (Dong et al., 2019). The interval of the basal divergence time was set to 5×10^6–6×10^6 (uniform), and the time difference between the first admixture event and the basal divergence time was set to 2.0×10^6–3.5×10^6. The large interval of 10^3–10^6 was used for the time difference between the two admixture events. We set 0–0.6 for both α (the proportion of the *G. metteniana* genome resulting from *G. gigantea* for the first admixture event) and β (the proportion of the *G. metteniana* genome resulting from *G. gigantea* for the second admixture event) according to our genetic admixture analysis. We set the large interval of 10^3–2.0×10^6 for the time difference between the polyploidy hybridization event and the time point when historical gene flow ended (models A1, A4, and A7) or recent gene flow began (models A2, A5, and A8). The interval of the recent population change was set to 5×10^5–2.0×10^6. The relative fit of the different demographic models to the data

was determined according to the Akaike information criterion and Akaike weights (Akaike 1974). The 95% confidence intervals for the parameters from the best model were calculated using a parametric bootstrapping approach with 50 independent runs for each bootstrap.

Ecological Niche Modeling and Niche Divergence

We used MAXENT v. 3.3.3 (Phillips and Dudík, 2008) to model the ecological niches of each species. We obtained the comprehensive current occurrence data of each species from both field collections and herbarium records. Data initially obtained from online databases were then double-checked to ensure minimal bias in the training set. To avoid the influence of autocorrelation among the locations used in the present study, we first removed sites that are too close (<10 km) to each other for each species before extracting values from the climate data sets, which resulted in a total of 55, 46, and 77 geo-referenced localities remained for *G. denticulata*, *G. metteniana*, and *G. gigantea*, respectively. Also, we checked "remove duplicate presence records" in the MAXENT program. Environmental layers of 19 bioclimatic variables (supplementary table S13, Supplementary Material online) for the present at a spatial resolution of 2.5 arc-minutes were downloaded from the WorldClim database (Hijmans et al. 2005). To exclude highly correlated climate variables, pairwise correlations were examined among the 19 variables within the distributional area for each species. Five bioclimatic variables with low correlation (Pearson's correlation < 0.7) were retained in our analysis: BIO2 (mean diurnal range, i.e., difference between maximum annual average temperature and minimum annual average temperature), BIO6 (minimum temperature of the coldest month), BIO7 (temperature annual range), BIO17 (precipitation of driest quarter), and BIO18 (precipitation of warmest quarter) (supplementary table S14, Supplementary Material online). The values of area under the receiver operating characteristic curve (AUC) were used to evaluate the accuracy of each model prediction. The threshold for good performance was set to 0.7 (Fielding and Bell 1997). The tenth percentile of training sample logistic threshold (suitability threshold that contained 90% of training samples) was used as a suitability threshold to divide high from low suitability areas for each species. Ecological niche modelings (ENMs) were generated for each species using ten crossvalidated replicate runs with the maximum number of background points of 10,000, a maximum of 5,000 iterations, and convergence threshold of 10^{-5}. A regularization multiplier of one was used because it developed a model with the highest AUC and lowest omission rate in our preliminary analysis (data not shown). The remaining parameters were set to default values.

We estimated the degree of niche overlap between each species with ENMTools v1.4.4 (Warren et al. 2010), based on the Schoener's *D* (Schoener 1968) and Warren's *I* statistics (Warren et al. 2008). For both parameters, a value of 0 means no overlap, whereas 1 denotes completely overlapping. We further performed the niche equivalency test, in order to test whether the ENMs of two species are identical (Warren et al. 2008). The null hypothesis of identical distribution is rejected when the observed scores of niche overlap statistics are significantly different from the values generated with 100 pseudoreplicates (Warren et al. 2008).

Supplementary Material

Supplementary data are available at Molecular Biology and Evolution online.

Acknowledgments

We would like to thank Yong Xiao and Huiqin Yi for their assistance in laboratory work and data analysis. This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDB31000000), Southeast Asia Biodiversity Research Institute, Chinese Academy of Science (Grant No. Y4ZK111B01 to M.K.), and the National Natural Science Foundation of China (Grant No. 31970218 to S.D.).

References

Akaike H. 1974. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 19(6):716–723.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3(10):e3376.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 16(1):37–48.

Barrera-Guzmán AO, Aleixo A, Shawkey MD, Weir JT. 2018. Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. *Proc Natl Acad Sci U S A*. 115(2):E218–E225.

Barrington DS. 1993. Ecological and historical factors in fern biogeography. *J Biogeogr*. 20(3):275–279.

Beck JB, Allison JR, Pryer KM, Windham MD. 2012. Identifying multiple origins of polyploid taxa: a multilocus study of the hybrid cloak fern (*Astrolepis integerrima*; Pteridaceae). *Am J Bot*. 99(11):1857–1865.

Bernabé N, Williams-Linera G, Palacios-Ríos M. 1999. Tree ferns in the interior and the edge of a Mexican cloud forest remnant: spore germination and sporophyte survival and establishment. *Biotropica* 31(1):83–88.

Caluff MG. 2002. A note on the Cuban tree fern hybrid *Cyathea* × *calolepis* (Cyatheaceae) and on its parentage. *Willdenowia* 32(2):311–318.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol*. 22(11):3124–3140.

Chung MY, Chung MG. 2013. Significant spatial aggregation and fine-scale genetic structure in the homosporous fern *Cyrtomium falcatum* (Dryopteridaceae). *New Phytol*. 199(3):663–672.

Clark J, Hidalgo O, Pellicer J, Liu HM, Marquardt J, Robert Y, Christenhusz M, Zhang SZ, Gibby M, Leitch IJ, et al. 2016. Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny. *New Phytol*. 210(3):1072–1082.

Comeault AA. 2018. The genomic and ecological context of hybridization affects the probability that symmetrical incompatibilities drive hybrid speciation. *Ecol Evol*. 8(5):2926–2937.

Conant DS. 1975. Hybrids in American Cyatheaceae. *Rhodora* 77:441–455.

Conant DS. 1978. A radioisotope technique to measure spore dispersal of the tree fern Cyathea arbores Sm. *Pollen Spores*. 20:583–593.

Conant DS. 1990. Observations on the reproductive biology of *Alsophila* species and hybrids (Cyatheaceae). *Ann Missouri Bot Gard*. 77(2):290–296.

Conant DS, Cooper-Driver G. 1980. Autogamous allohomoploidy in *Alsophila and Nephelea*. *Am J Bot*. 67(9):1269–1288.

Coyne JA, Orr HA. 2004. Speciation. Sunderland (MA): Sinauer Associates.
Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.
Dauphin B, Grant JR, Farrar DR, Rothfels CJ. 2018. Rapid allopolyploid radiation of moonwort ferns (*Botrychium*; Ophioglossaceae) revealed by PacBio sequencing of homologous and homeologous nuclear regions. *Mol Phylogenet Evol*. 120:342–353.
Dong SY, Xiao Y, Kong H, Feng C, Harris AJ, Yan Y, Kang M. 2019. Nuclear loci developed from multiple transcriptomes yield high resolution in phylogeny of scaly tree ferns (Cyatheaceae) from China and Vietnam. *Mol Phylogenet Evol*. 139:106567.
Dong SY, Zuo ZY. 2018. On the recognition of *Gymnosphaera* as a distinct genus in Cyatheaceae. *Ann Missouri Bot Gard*. 103(1):1–23.
Douglas GM, Gos G, Steige KA, Salcedo A, Holm K, Josephs EB, Arunkumar R, Ågren JA, Hazzouri KM, Wang W, et al. 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proc Natl Acad Sci U S A*. 112(9):2806–2811.
Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*. 19:11–15.
Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annu Rev Genet*. 42(1):443–461.
Earl DA, Vonholdt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Cons Genet Resour*. 4(2):359–361.
Edwards PJ. 2005. Cyatheaceae. In Beentje HJ and Ghazanfar SA, editors. Flora of tropical East Africa. Richmond, Surrey (UK): Royal Botanic Gardens, Kew. p. 1–15.
Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol*. 14(8):2611–2620.
Excoffier L, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet*. 9(10):e1003905.
Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 10(3):564–567.
Fielding AH, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Envir Conserv*. 24(1):38–49.
Flot JF. 2010. SeqPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Mol Ecol Resour*. 10(1):162–166.
Fowler NL, Levin DA. 2016. Critical factors in the establishment of allopolyploids. *Am J Bot*. 103(7):1236–1251.
Fujiwara T, Serizawa S, Watano Y. 2018. Phylogenetic analysis reveals the origins of tetraploid and hexaploid species in the Japanese *Lepisorus thunbergianus* (Polypodiaceae) complex. *J Plant Res*. 131(6):945–959.
Gastony GJ, Yatskievych. 1992. Maternal inheritance of the chloroplast and mitochondrial genomes in Cheilanthoid ferns. *Am J Bot*. 79(6):716–722.
Grusz AL, Rothfels CJ, Schuettpelz E. 2016. Transcriptome sequencing reveals genome-wide variation in molecular evolutionary rate among ferns. *BMC Genomics*. 17(1):692.
Grusz AL, Windham MD, Pryer KM. 2009. Deciphering the origins of apomictic polyploids in the *Cheilanthes yavapensis* complex (Pteridaceae). *Am J Bot*. 96(9):1636–1645.
Han TS, Wu Q, Hou XH, Li ZW, Zou YP, Ge S, Guo YL. 2015. Frequent introgressions from diploid species contribute to the adaptation of the tetraploid shepherd's purse (*Capsella bursa-pastoris*). *Mol Plant*. 8(3):427–438.
Haufler C. 2002. Homospory: an odyssey of progress in pteridophyte genetics and evolutionary biology. *BioScience* 52(12):1081–1093.
Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol*. 25(15):1965–1978.
Hunt HV, Ansell SW, Russell SJ, Schneider H, Vogel JC. 2011. Dynamics of polyploid formation and establishment in the allotetraploid rock fern *Asplenium majoricum*. *Ann Bot*. 108(1):143–157.
Husband BC. 2004. The role of triploid hybrids in the evolutionary dynamics of mix-ploidy populations. *Biol J Linn Soc Lond*. 82(4):537–546.
Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 23(2):254–267.
Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 23(14):1801–1806.
Janssen T, Bystriakova N, Rakotodranibe F, Coomes D, Labat JN, Schneider H. 2008. Neoendemism in Madagascan scaly tree ferns results from recent, coincident diversification bursts. *Evolution* 62(8):1876–1889.
Janssen T, Rakotondrainibe F. 2007. An update of the revision of *Cyathea* subgen. *Alsophila* sect. *Gymnosphaera* (Cyatheaceae) in Madagascar and the Comoros including a discussion of putative hybridization events. *Adansonia* 29:195–213.
Jørgensen MH, Ehrich D, Schmickl R, Koch MA, Brysting AK. 2011. Interspecific and interploidal gene flow in Central European *Arabidopsis* (Brassicaceae). *BMC Evol Biol*. 11(1):346–359.
Kawakami SM, Kondo K, Kawakami S. 2010. Reticulate evolution of the hybrid produced artificially by crosses between *Osmunda banksiifolia* and *Osmunda lancea*. *J Plant Res*. 123(5):639–644.
Kong HH, Condamine FL, Harris AJ, Chen JL, Pan B, Möller M, Hoang VS, Kang M. 2017. Both temperature fluctuations and East Asian monsoons have driven plant diversification in the karst ecosystems from southern China. *Mol Ecol*. 26(22):6414–6429.
Korall P, Conant DS, Metzgar JS, Schneider H, Pryer KM. 2007. A molecular phylogeny of scaly tree ferns (Cyatheaceae). *Am J Bot*. 94(5):873–886.
Korall P, Pryer KA, Metzgar JS, Schneider H, Conant DS. 2006. Tree ferns: monophyletic groups and their relationships as revealed by four protein-coding plastid loci. *Mol Phylogenet Evol*. 39(3):830–845.
Korall P, Pryer KM. 2014. Global biogeography of scaly tree ferns (Cyatheaceae): evidence for Gondwanan vicariance and limited transoceanic dispersal. *J Biogeogr*. 41(2):402–413.
Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR. 2008. The ups and downs of genome size evolution in polyploid species of nicotiana (Solanaceae). *Ann Bot*. 101(6):805–814.
Li Y, Zhang YL, Jiang CD, Wang T, Wang Q, Shi L. 2010. Effect of storage temperature on spore viability and early gametophyte development of three vulnerable species of *Alsophila* (Cyatheaceae). *Aust J Bot*. 58(2):89–96.
Luo X, Hu Q, Zhou P, Zhang D, Wang Q, Abbott RJ, Liu J. 2017. Chasing ghosts: allopolyploid origin of *Oxyria sinensis* (Polygonaceae) from its only diploid congener and an unknown ancestor. *Mol Ecol*. 26(11):3037–3049.
Ma JX, Li YN, Vogl C, Ehrendorfer F, Guo YP. 2010. Allopolyploid speciation and ongoing backcrossing between diploid progenitor and tetraploid progeny lineages in the *Achillea millefolium* species complex: analyses of single-copy nuclear genes and genomic AFLP. *BMC Evol Biol*. 10(1):100.
Menken S, Smit E, Den Nijs H. 1995. Genetical population structure in plants: gene flow between diploid sexual and triploid asexual dandelions (*Taraxacum* Section *Ruderalia*). *Evolution* 49(6):1108–1118.
Miller MA, Schwartz T, Pickett BE, He S, Klem EB, Scheuermann RH, Passarotti M, Kaufman S, O'Leary MA. 2015. A RESTful API for access to phylogenetic tools via the CIPRES science gateway. *Evol Bioinform Online*. 11:43–48.
Nakato N. 1989. Cytological studies on the genus *Cyathea* in Japan. *J Jap Bot*. 64:142–147.
Nakazato T, Jung MK, Housworth EA, Rieseberg LH, Gastony GJ. 2007. A genomewide study of reproductive barriers between allopatric populations of a homosporous fern, *Ceratopteris richardii*. *Genetics* 177(2):1141–1150.
Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Parisod C. 2012. Polyploids integrate genomic changes and ecological shifts. *New Phytol*. 193(2):297–300.
Payseur BA, Rieseberg LH. 2016. A genomic perspective on hybridization and speciation. *Mol Ecol*. 25(11):2337–2360.
Perrie LR, Brownsey PJ. 2005. Insights into the biogeography and polyploid evolution of New Zealand *Asplenium* from chloroplast DNA sequence data. *Am Fern J*. 95(1):1–21.
Perrie LR, Shepherd LD, De Lange PJ, Brownsey PJ. 2010. Parallel polyploid speciation: distinct sympatric gene-pools of recurrently derived allo-octoploid *Asplenium* ferns. *Mol Ecol*. 19(14):2916–2932.
Petit C, Bretagnolle F, Felber F. 1999. Evolutionary consequences of diploid-polyploid hybrid zones in wild species. *Trends Ecol Evol*. 14(8):306–311.
Phillips SJ, Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31(2):161–175.
Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14(9):817–818.
PPG I. 2016. A community-derived classification for extant lycophytes and ferns. *J Syst Evol*. 54:563–603.
Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959.
R Core Development Team. 2010. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.R-project.org/
Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst Biol*. 67(5):901–904.
Ramírez-Barahona S, Eguiarte LE. 2015. Spatial genetic analyses reveal strong genetic structure in two populations of the outcrossing tree fern *Alsophila firma* (Cyatheaceae). *Bot J Linn Soc*. 177(3):439–449.
Ramírez-Barahona S, Luna-Vega I, Tejero-Díez D. 2011. Species richness, endemism, and conservation of American tree ferns (Cyatheales). *Biodivers Conserv*. 20(1):59–72.
Ramsey J. 2011. Polyploidy and ecological adaptation in wild yarrow. *Proc Natl Acad Sci U S A*. 108(17):7096–7101.
Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu Rev Ecol Syst*. 29(1):467–501.
Ramsey J, Schemske DW. 2002. Neopolyploidy in flowering plants. *Annu Rev Ecol Syst*. 33(1):589–639.
Ranker TA, Sundue MA. 2015. Why are there so few species of ferns? *Trends Plant Sci*. 20(7):402–403.
Rice A, Glick L, Abadi S, Einhorn M, Kopelman NM, Salman-Minkov A, Mayzel J, Chay O, Mayrose I. 2015. The Chromosome Counts Database (CCDB): a community resource of plant chromosome numbers. *New Phytol*. 206(1):19–26.
Rieseberg LH, Willis JH. 2007. Plant speciation. *Science* 317(5840):910–914.
Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 61(3):539–542.
Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes*. 4(1):137–138.
Rothfels CJ, Johnson AK, Windham MD, Pryer KM. 2014. Low-copy nuclear data confirm rampant allopolyploidy in the Cystopteridaceae (Polypodiales). *Taxon* 63(5):1026–1036.
Roux C, Pannell JR. 2015. Inferring the mode of origin of polyploid species from next-generation sequence data. *Mol Ecol*. 24(5):1047–1059.
Schneider H, Liu H-M, Chang Y-F, Ohlsen D, Perrie LR, Shepherd L, Kessler M, Karger DN, Hennequin S, Marquardt J, et al. 2017. Neo-and Paleopolyploidy contribute to the species diversity of asplenium: the most species-rich genus of ferns. *J Syst Evol*. 55(4):353–364.
Schoener TW. 1968. The Anolis lizards of Bimini: resource partitioning in a complex fauna. *Ecology* 49(4):704–726.
Schumer M, Cui R, Rosenthal G, Andolfatto P. 2015. Reproductive isolation of hybrid populations driven by genetic incompatibilities. *PLoS Genet*. 11(3):e1005041.
Sessa EB, Zimmer EA, Givnish TJ. 2012. Reticulate evolution on a global scale: a nuclear phylogeny for New World *Dryopteris* (Dryopteridaceae). *Mol Phylogenet Evol*. 64(3):563–581.
Sigel EM. 2016. Genetic and genomic aspects of hybridization in ferns. *J Syst Evol*. 54(6):638–655.
Sigel EM, Windham MD, Pryer KM. 2014. Evidence for reciprocal origins in *Polypodium hesperium* (Polypodiaceae): a fern model system for investigating how multiple origins shape allopolyploid genomes. *Am J Bot*. 101(9):1476–1485.
Slotte T, Huang H, Lascoux M, Ceplitis A. 2008. Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Brassicaceae). *Mol Bio Evol*. 25(7):1472–1481.
Soltis DE, Visger CJ, Soltis PS. 2014. The polyploidy revolution then. . .and now: Stebbins revisited. *Am J Bot*. 101(7):1057–1078.
Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annu Rev Plant Biol*. 60(1):561–588.
Sora D, Kron P, Husband BC. 2016. Genetic and environmental determinants of unreduced gamete production in *Brassica napus*, *Sinapis arvensis* and their hybrids. *Heredity* 117(6):440–448.
Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 68(4):978–989.
Sutherland BL, Galloway LF. 2017. Postzygotic isolation varies by ploidy level within a polyploid complex. *New Phytol*. 213(1):404–412.
Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony (* and other methods). Version 4.0b10. Sunderland (MA): Sinauer Associates.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123(3):585–595.
Testo WL, Watkins JE, Barrington DS. 2015. Dynamics of asymmetrical hybridization in North American wood ferns: reconciling patterns of inheritance with gametophyte reproductive biology. *New Phytol*. 206(2):785–795.
Tryon RM, Gastony GJ. 1975. The biogeography of endemism in the Cyatheaceae. *Fern Gaz*. 11:73–79.
Vallejo-Marín M, Buggs RJA, Cooley AM, Puzey JR. 2015. Speciation by genome duplication: repeated origins and genomic composition of the recently formed allopolyploid species *Mimulus peregrinus*. *Evolution* 69(6):1487–1500.
Wagner WH, Wagner FS, Reznicek AA, Werth CR. 1992. ×*Dryostichum singulare* (Dryopteridaceae), a new fern nothogenus from Ontario. *Can J Bot*. 70(2):245–253.
Walker S. 1961. Cytogenetic studies in the *Dryopteris spinulosa* complex II. *Am J Bot*. 48(7):607–614.
Wan S, Li AC, Clift PD, Stuut J. 2007. Development of the East Asian monsoon: mineralogical and sedimentological records in the northern South China sea since 20 Ma. *Palaeogeogr Palaeoclimatol Palaeoecol*. 254(3-–4):561–582.
Wang L, Schneider H, Zhang XC, Xiang QP. 2012. The rise of the Himalaya enforced the diversification of SE Asian ferns by altering the monsoon regimes. *BMC Plant Biol*. 12(1):210.
Warren DL, Glor RE, Turelli M. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62(11):2868–2883.
Warren DL, Glor RE, Turelli M. 2010. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* 33:607–611.
Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci U S A*. 106(33):13875–13879.
Wright S. 1984. Evolution and the genetics of populations. Chicago (IL): University of Chicago Press.
Xiang L, Werth CR, Emery SN, McCauley DE. 2000. Population-specific gender-biased hybridization between *Dryopteris intermedia* and

D. carthusiana: evidence from chloroplast DNA. *Am J Bot*. 87(8):1175–1180.

Zhang R, Liu T, Wu W, Li Y, Chao L, Huang L, Huang Y, Shi S, Zhou R. 2013. Molecular evidence for natural hybridization in the mangrove fern genus *Acrostichum*. *BMC Plant Biol*. 13(1):74.

Zhang XC, Nishida H. 2013. Cyatheaceae. In: ZY Wu, PH Raven, DY Hong, editors. Flora of China. Vols. 2–3. Beijing (China): Science Press. p. 134–138.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28(24):3326–3328.

Zohren J, Wang N, Kardailsky I, Borrell JS, Joecker A, Nichols RA, Buggs RJ. 2016. Unidirectional diploid-tetraploid introgression among British birch trees with shifting ranges shown by RAD markers. *Mol Ecol*. 25(11):2413–2426.