



OPEN

The dynamic evolution of mobile open reading frames in plastomes of *Hymenophyllum* Sm. and new insight on *Hymenophyllum coreanum* Nakai

Hyung Tae Kim¹ & Jung Sung Kim²✉

In this study, four plastomes of *Hymenophyllum*, distributed in the Korean peninsula, were newly sequenced and phylogenomic analysis was conducted to reveal (1) the evolutionary history of plastomes of early-diverging fern species at the species level, (2) the importance of mobile open reading frames in the genus, and (3) plastome sequence divergence providing support for *H. coreanum* to be recognized as an independent species distinct from *H. polyanthos*. In addition, 1C-values of *H. polyanthos* and *H. coreanum* were measured to compare the genome size of both species and to confirm the diversification between them. The *rrn16-trnV* intergenic regions in the genus varied in length caused by Mobile Open Reading Frames in Fern Organelles (MORFFO). We investigated enlarged noncoding regions containing MORFFO throughout the fern plastomes and found that they were strongly associated with tRNA genes or palindromic elements. Sequence identity between plastomes of *H. polyanthos* and *H. coreanum* is quite low at 93.35% in the whole sequence and 98.13% even if the variation in *trnV-rrn16* intergenic spacer was ignored. In addition, different genome sizes were found for these species based on the 1C-value. Consequently, there is no reason to consider them as a conspecies.

Plastomes of vascular plants are varied in size (120–170 kb) and compact with 110–130 genes^{1–5}. In angiosperms, they are normally conserved in sequence, gene contents and organization¹ except in heterotrophic plants^{6,7} or specific lineages^{8–10}. Consequently, single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) among infraspecific taxa or individuals in angiosperms are very low^{11–14}. In contrast to angiosperms, the genetic variation in plastomes between individuals in ferns is almost similar to the interspecific genetic variation in angiosperms (e.g., for *Equisetum arvense*¹⁵). This is likely due to the longer evolutionary history of ferns than angiosperms.

Since the first plastome sequence of *Adiantum capillus-veneris* L. was reported¹⁶, plastomes from 11 orders of extant ferns¹⁷ have also been sequenced^{18–22}. These data have given insights into (1) phylogenomics for resolving the deep relationships throughout fern lineages^{19,23,24}, (2) inserted foreign DNA^{3,25}, and (3) lineage-specific structural evolutions^{3,26,27}. However, because a large proportion of plastome sequence data in ferns belongs to the order Polypodiales, which is the most derived order in the extant ferns¹⁷, the evolution of plastomes at a low taxonomic level in eusporangiate ferns and basal leptosporangiate ferns is still equivocal.

Hymenophyllaceae, filmy ferns, is a basal family of leptosporangiate ferns¹⁷, and it consists of more than 600 species²⁸. Their distinctive feature, single-cell thick laminae, easily distinguishes them from other fern families²⁹. However, infra-familial classification of the filmy ferns has been argued for a long time^{30–33}. Nevertheless, Hymenophyllaceae is traditionally classified into two major clades, “trichomanoid” with obconic or tubular involucre and “hymenophylloid” with bivalvate involucre. This relationship has also been clarified using recent molecular analysis²⁸.

¹Institute of Agricultural Science and Technology, Chungbuk National University, Cheongju, Chungbuk 28644, Korea. ²Department of Forest Science, Chungbuk National University, Cheongju, Chungbuk 28644, Korea. ✉email: jungsung@chungbuk.ac.kr

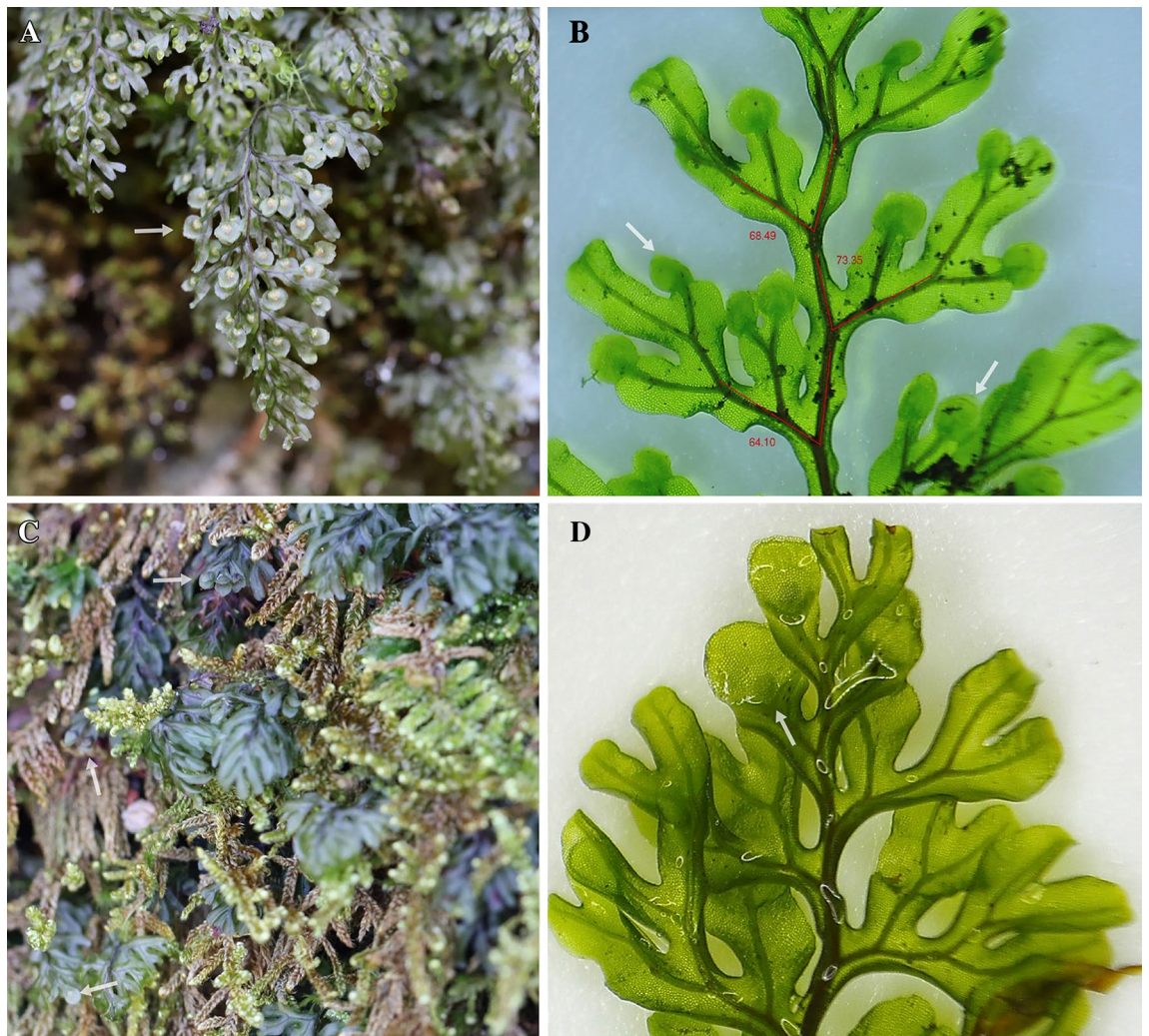


Figure 1. Phenotypes of *Hymenophyllum polyanthos* (A and B) and *H. coreanum* (C and D). (A) Photo of a whole plant of *H. polyanthos* in the field. (B) Microphotograph of *H. polyanthos*. (C) *H. coreanum* in the field. (D) Microphotograph of *H. coreanum*. White arrow refers to the involucre and red angle refers to angle from rachis to pinna.

Hymenophyllum s.l. is comprised of more than 300 species with a nearly cosmopolitan distribution³⁴. Pryer et al.²⁸ showed that *Serpyllopsis*, *Cardiomanes*, and *Microtrichmanes*, which were segregated from each other or belonged to *Trichomanes* by transitional classifications^{30,32,33,35}, were included within *Hymenophyllum s.l.* Although, they also mentioned that their data set was insufficient in terms of taxon sampling and the number of genes used to evaluate all of the segregates within *Hymenophyllum s.l.* Hennequin et al.³⁶ reconstructed the phylogenetic tree of *Hymenophyllum s.l.* using increased sampling and additional genes and divided *Hymenophyllum s.l.* into eight subgenera. Interestingly, four segregate genera and one section belonging to *Trichomanes* formed a clade with *Hymenophyllum* species, in agreement with Pryer et al.²⁸. Recently, Ebihara et al.²⁹ divided *Hymenophyllum s.l.* into 10 subgenera based on molecular phylogenetic analyses and macroscopic characters.

In Korea, three species and one unacceptable taxon of the genus *Hymenophyllum* have been reported: *Hymenophyllum barbatum* (Bosch) Baker, *H. polyanthos* (Sw.) Sw., *H. wrightii* Bosch and *H. coreanum* Nakai. Among them, *H. coreanum* was first recognized by Nakai³⁷ who collected the sample from Kumgangsán in Gangwondo, Korea and described it as a plant with 2–5 mm stipes, 5–15 mm fronds and dense sori in apex fronds. In contrast to *H. coreanum*, *H. polyanthos* in Korea is 1.5–3 times longer and has involucres from the apex to the middle of the frond (Fig. 1)³⁸. However, in spite of these different characteristics, *H. coreanum* has been considered a synonym of *H. polyanthos* by many researchers. The broadly described species, *H. polyanthos* has a worldwide distribution but probably includes a number of distant lineages in the subgenus *Mecodium* that do not have any specialized morphological characters³⁹. The recent phylogeny of *Hymenophyllum s.l.* using molecular markers revealed that *H. polyanthos* was highly polyphyletic^{36,40}. Therefore, more clear scientific evidence for considering the correct taxonomical position of *H. coreanum* and its relationship with *H. polyanthos* is necessary.

In this study, we sequenced four new plastomes of *Hymenophyllum* species distributed in the Korean Peninsula: (1) to explore the plastome evolution of early-diverging leptosporangiate ferns at the species level, (2)

Subgenus	Species	Total length (bp)	LSC (bp)	SSC (bp)	IR (bp)	GC content (%)	Coverage depth (X)
<i>Hymenophyllum</i>	<i>H. barbatum</i>	144,112	102,259	20,789	10,532	37.6	227.9
	<i>H. holochilum</i> ^a	142,214	99,299	20,765	11,075	37.6	
<i>Mecodium</i>	<i>H. coreanum</i>	157,967	89,616	21,277	23,537	38.2	367.3
	<i>H. polyanthos</i>	160,865	89,811	21,290	24,882	38.6	313.7
	<i>H. wrightii</i>	149,309	90,003	21,266	19,020	37.5	325.3

Table 1. Summary of plastome sequences of *Hymenophyllum*. ^aDownloaded from Kuo et al. (2018).

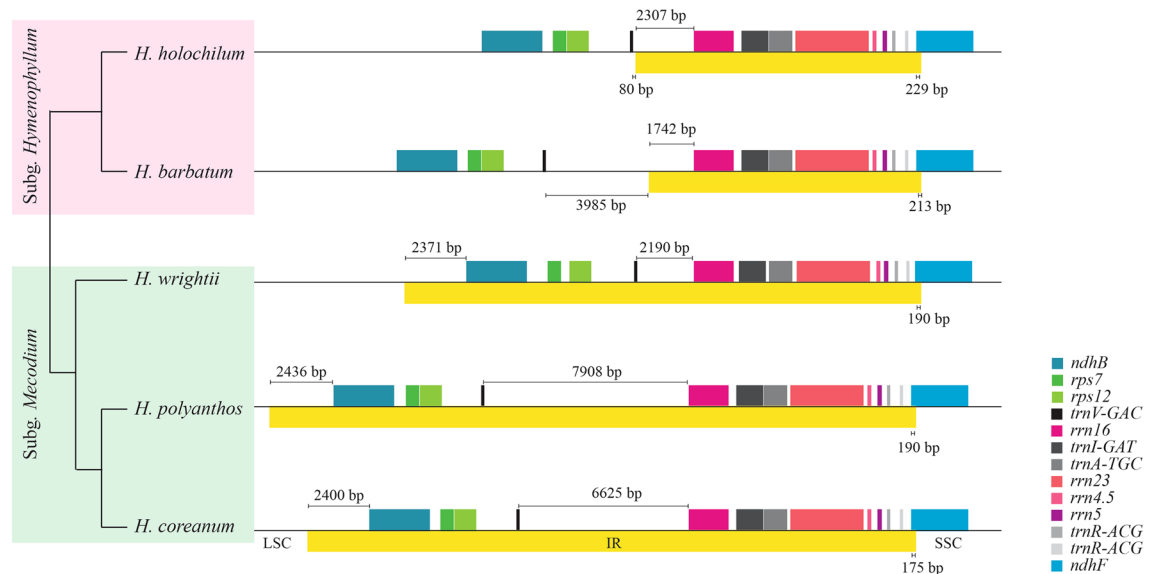


Figure 2. IR boundaries in the plastomes of *Hymenophyllum*. The phylogeny of *Hymenophyllum* species is extracted from the phylogeny of Hymenophyllaceae using 85 genes in this study. The boxes above the black line refer to the genes and the yellow box below the line refers to the region.

to verify the Mobile Open Reading Frames in Fern Organelles (MORFFO) which were identified by Robison et al.²⁵, and (3) to investigate the possibility that the two morphologically similar species, *H. polyanthos* and *H. coreanum*, are distinct species. In addition, 1C-values of both species were measured to compare genome size and to explore the diversification between *H. polyanthos* and *H. coreanum*.

Results

Genome structure and gene contents of plastomes in *Hymenophyllum*. The newly sequenced plastomes of the four *Hymenophyllum* species were from 144,112 bp to 160,865 bp in length with a GC content of 37.5–38.6% (Table 1). The large single copy (LSC) region in the subgenus *Hymenophyllum* was 10 kb longer than that in the subgenus *Mecodium*, due to inverted repeat (IR) expansion of the LSC region (Fig. 2). The length difference of the small single copy (SSC) region between the two subgenera mainly resulted from the insertions at the *rpl32-trnP* and *ndhA* introns in the subgenus *Mecodium*. The IR-SSC boundary was slightly different among species but positioned near the 5' end of *ndhF*. There were 85 coding genes, 8 rRNA genes, 33 tRNA genes, and one pseudogene (*trnL-CAA*) in the subgenus *Hymenophyllum*. However, three species in the subgenus *Mecodium* had one more *rps7*, *ndhB* and *trnV* due to IR expansion. Gene order remained stable among species in the genus *Hymenophyllum* except the IR expansion of the subgenus *Mecodium*.

Length variation hotspots in *Hymenophyllum* plastomes. The intergenic spacer (IGS) between *trnV* and *rrn16* was the most variable region in length (Fig. 2). Because coverage depths between three plant genomes in a NGS data were significantly different⁴¹, stable read depths throughout the plastomes in this study confirmed that these length variations did not result from mitochondrial plastome sequences or nuclear plastome sequences.

In subgenus *Hymenophyllum*, *trnV-rrn16* of *H. holochilum* was 2387 bp in length and mostly belonged to the IR region. However, that of *H. barbatum* was expanded 3.4 kb more than *H. holochilum* and two-thirds of this region was part of the LSC region (Supplementary Table 1). In the subgenus *Mecodium*, *trnV-rrn16* of *H. wrightii* was 2190 bp in length and 5.7 kb and 4.4 kb expansions were confirmed at the plastomes of *H. polyanthos* and *H. coreanum*, respectively. In total, 20 ORFs having more than 100 amino acid (aa) sequences in length were also found in this length-variable region of five *Hymenophyllum* species (Supplementary Fig. 1). These ORFs ranged

Query sequence			Blastp result					
Sequence Name	Name	Length (aa ^a)	Organism	Gene	Query cover (%)	E-value	Identity	Accession
<i>H. holochilum</i>	ORF_H1	103	<i>Pinus koraiensis</i>	ORF46h	20	3.40E-02	85.71	YP_001152247.1
	ORF_H2	363	<i>Mankyua chejuense</i>	ORF295	73	3.00E-76	46.99	YP_005352949.1
			<i>Roya anglica</i>	hypothetical protein	44	3.00E-07	26.95	YP_009033761.1
	ORF_H3	150	- ^b					
<i>H. barbatum</i>	ORF_B1	1,274	<i>Mankyua chejuense</i>	ORF531	41	4.00E-130	42.45	ADZ47985.2
			<i>Mankyua chejuense</i>	ORF295	22	3.00E-84	48.12	YP_005352949.1
			<i>Beggiatoa leptomitiformis</i>	AL038_02335	33	2.00E-24	28.01	ALG66761.1
	ORF_B2	101	-					
ORF_B3	280	-						
<i>H. wrightii</i>	ORF_W1	216	-					
	ORF_W2	132	-					
<i>H. polyanthos</i>	ORF_P1	664	<i>Mankyua chejuense</i>	ORF295	44	3.00E-86	48.12	YP_005352949.1
			<i>Mankyua chejuense</i>	ORF187	18	3.00E-22	41.94	YP_005352954.1
	ORF_P2	222	-					
	ORF_P3	107	-					
	ORF_P4	147	<i>Mankyua chejuense</i>	ORF531	97	9.00E-14	37.67	YP_005352953.1
			<i>Bacteroidetes bacterium</i>	hypothetical protein	59	1.50E-01	33.33	HAV23789.1
	ORF_P5	572	<i>Mankyua chejuense</i>	ORF295	52	5.00E-77	45.97	YP_005352949.1
	ORF_P6	129	<i>Mankyua chejuense</i>	ORF531	82	5.00E-14	37.04	YP_005352953.1
ORF_P7	107	<i>Thiobaca trueperi</i>	peptide chain release factor 3	58	3.10E-01	43.75	WP_132977102.1	
<i>H. coreanum</i>	ORF_C1	1,064	<i>Mankyua chejuense</i>	ORF531	47	1.00E-123	43.24	YP_005352953.1
			<i>Beggiatoa leptomitiformis</i>	hypothetical protein	46	4.00E-25	25.54	WP_083991398.1
			<i>Thiotrichales bacterium HS_08</i>	hypothetical protein	43	3.00E-18	24.7	WP_103918394.1
	ORF_C2	218	-					
	ORF_C3	359	-					
	ORF_C4	133	-					
	ORF_C5	200	-					

Table 2. Blastp results for ORFs within *trnV-rrn16* in *Hymenophyllum* plastomes. ^aAmino acid. ^bNo significant similarity found.

from 101 aa to 1274 aa in length and blastp search showed that half of them had similar amino acid sequences to ORFs that were embedded in the plastome of *Mankyua chejuense*³ (Table 2). Based on the DNA sequences, *trnV-rrn16* of five *Hymenophyllum* species contained at least a part of MORFFO. In addition, *H. wrightii* and *H. holochilum* contained a part of MORFFO at *trnE-trnG* and *rrn16-ycf2*, respectively.

Enlarged noncoding regions in fern plastomes. To investigate the relationship between MORFFO and inserted loci, expanded noncoding regions having a part of MORFFO in available fern plastomes (138 plastomes, Supplementary Table 2) were found using blastn search. In total, 32 loci were matched to MORFFO with at least 100 bp in length (Supplementary table 3). Among them, 22 loci (75%) were flanked by tRNA genes. The most frequent locus, including MORFFO, throughout the plastomes was *rrn16-(trnV-GAC)-rps12* in which *trnV-GAC* was generally intact in eusporangiate ferns and early-diverging leptosporangiate ferns but was deleted or pseudogenized in most core leptosporangiate ferns.

In addition, all expanded regions had palindrome sequences at both ends with a minimum stem length of 5 bp, a minimum loop length of 5 bp, and a minimum stem-loop sequence of 20 bp.

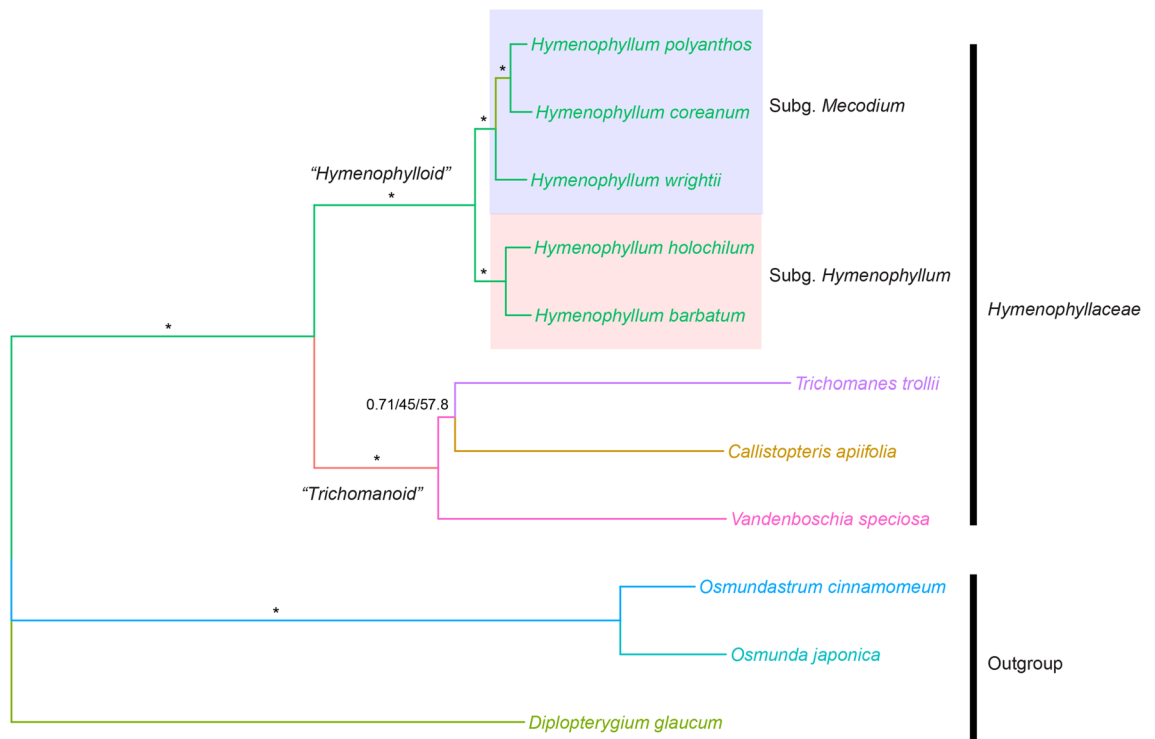


Figure 3. Phylogeny of the family Hymenophyllaceae using Bayesian inference with 85 genes. Numbers on the branches refer to Bayesian posterior probability/bootstraps support of ML/bootstraps support of MP. *On the branches stands for the supporting values of 1/100/100.

Phylogenetic relationships of *Hymenophyllum* species. The concatenated alignment using 85 coding genes was 74,933 bp in length. Of these characters, 45,301 were constant, 11,129 variable characters were parsimonious uninformative, and 18,503 were parsimonious informative. Tree topologies among maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference were completely identical with maximum support on all branches, with the exception of the moderate support on the clade *Callistopteris apiifolia* + *Trichomanes trollii* (Fig. 3). Hymenophyllaceae species were monophyletic and two traditional major clades, *Hymenophylloid* and *Trichomanoid*, were also strongly supported. *H. coreanum* was sister to *H. polyanthos* and they formed a clade with *H. wrightii*. As a result, two subgenera in *Hymenophyllum* were well resolved.

Genome size variation between *H. polyanthos* and *H. coreanum*. Because living materials of *H. coreanum* and *H. polyanthos* were very restricted in Korea, the genome size measurements for both taxa were carried out with two populations from two distant collecting sites and three populations from two distant sites, respectively. The measurement was repeated over three times for each sample. Compared to the calibration standard (1C-value of *N. tabacum* = 5.18⁴²), the 1C-values of *H. polyanthos* and *H. coreanum* were 16.16 ± 0.17 and 14.85 ± 0, respectively (Fig. 4).

Discussion

Taxonomical position of *H. coreanum* and relationship with *H. polyanthos*. The genus *Hymenophyllum*, delimited by Morton³⁰ and Iwatsuki³⁵, was recently reclassified into 10 subgenera by Ebihara et al.²⁹ based on morphological characters and molecular evidence³⁶. Among the 10 subgenera, sensu Ebihara et al.²⁹, within *Hymenophyllum*, eight subgenera have the basic chromosome number of $x = 36$. However, the subgenus *Mecodium* has a chromosome number of $x = 28$ and the subgenus *Hymenophyllum* has chromosome numbers of $x = 11$ to 28²⁹. Because the latter two subgenera seemed to be the most recently derived taxa within the genus³⁶, the ancestor of Hymenophyllales was suggested to have the chromosome number $2n = 72$ ⁴³, and it is assumed that the reduction in chromosome number is a synapomorphy of these two subgenera in the genus *Hymenophyllum*⁴⁴.

According to Ebihara et al.²⁹, the subgenus *Mecodium* consisted of *H. polyanthos* and its local derivatives. On the other hand, it was noted that *H. polyanthos* shows variation in its chromosome number^{45,46} ($n = 28$ Japanese taxa or 27 taxa of other country outside of Japan), and more recently, the polyphyly of this cosmopolitan species, including several regional taxa, was discovered^{36,40}. Therefore, the $n = 27$ population of *H. polyanthos* mentioned by Iwatsuki⁴⁶ is likely to be the result of a transition from $n = 28$ to $n = 27$.

In this study, the chromosome numbers of *H. polyanthos* and *H. coreanum* were not directly counted. However, the result of flow cytometry implied that the genome of *H. coreanum* is downsized compared with that of *H. polyanthos*. It is not clear whether *H. coreanum* has been adopted to the $n = 27$ population of *H. polyanthos*,

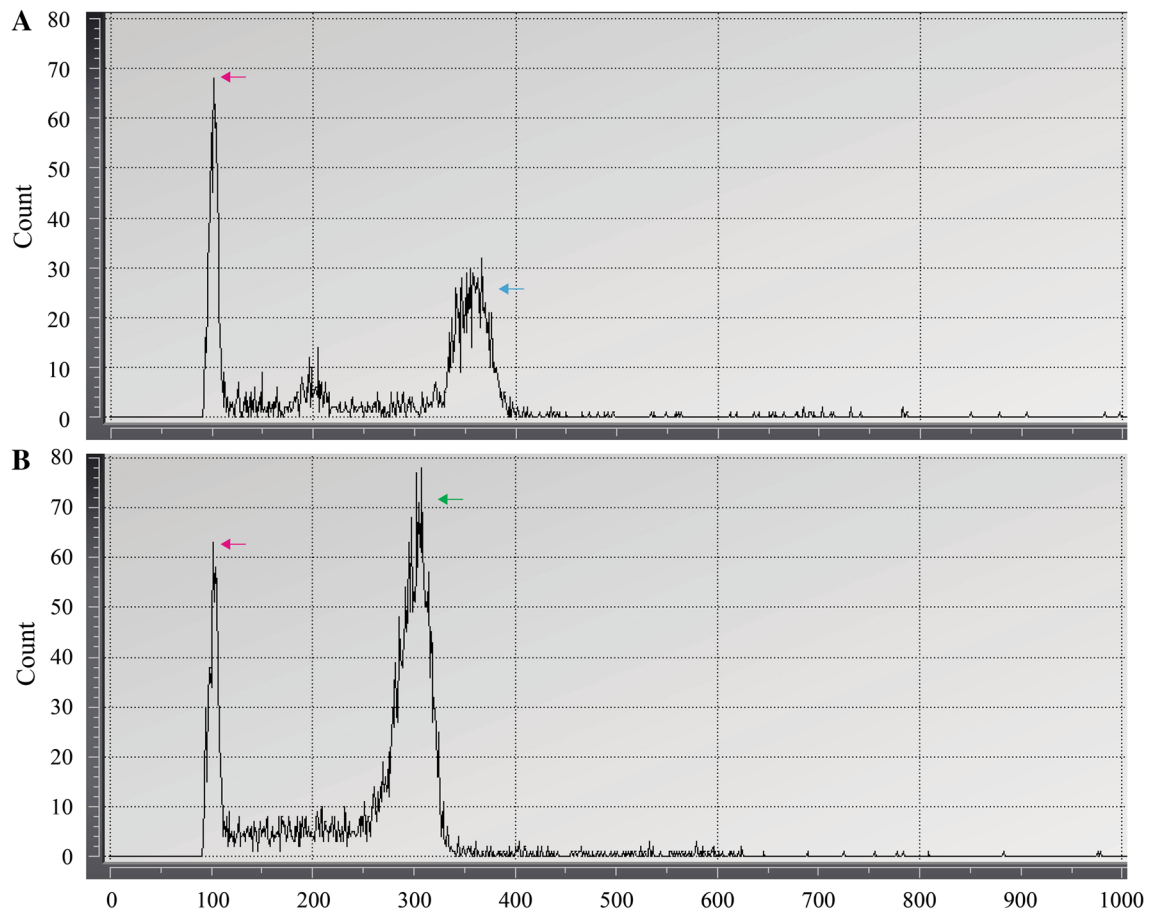


Figure 4. The results of the genome size measurement using the flow cytometer. (A) *H. polyanthos* (Blue arrow) (B) *H. coreanum* (Green arrow). Red arrow refers to *Nicotiana tabacum* as reference.

mentioned by Iwatsuki⁴⁶, based on the data at hand. However, it is clear that *H. coreanum* is distinct from *H. polyanthos* based on morphological characteristics and genome size. Therefore, *H. coreanum* should be considered as an independent species not one of the synonymous taxa of *H. polyanthos*.

The divergence of plastome sequences between both species also strongly supports this taxonomical treatment. So far, it was rare to report more than two plastome sequences in the same species of ferns (Fig. 5 and supplementary table 4). The percentage of sequence identity between two plastomes in a species ranges from 98.86 (*Equisetum arvense*) to 99.97% (*Mankyua chejuense*). Except for the case of *E. arvense* in which the copy number variations frequently occurred¹⁵, sequence identities within the same species are higher than 99.7% in ferns. In contrast, the sequence identities between the plastomes of two different species in a genus, with the exception of the genera *Cyrtomium* and *Azolla*, are less than 98.7% (Fig. 5 and supplementary table 4). Sequence identity between plastomes of *H. polyanthos* and *H. coreanum* is quite low at 93.35% in the whole sequence and 98.13% even if the variation in *trnV-rrn16* IGS was ignored. Consequently, plastome sequences of these two species revealed that they are quite distinct from each other and there is no reason to consider them as an intraspecific variation.

In this study, we only investigated the Korean population of *H. coreanum* and it is not known whether the *H. coreanum* distributed in China, Taiwan and Japan³⁸ also has the same genome size and plastome sequences as the Korean population. In Korea, the species mainly inhabits high mountains, however information for this species is very poor because it has previously been considered a synonymous taxon of *H. polyanthos*. Further study of this species facilitate better understanding of the speciation of the subgenus *Mecodium* because many species within the subgenus are derivatives of *H. polyanthos*, and *H. coreanum*, at least in Korea, seems to be recently diverged from *H. polyanthos*.

MORFFO of ferns prefer stem-loop structures. In previous studies^{25,47}, the physical positions of MORFFO were considered to be related to inversions or IR borders even though it is not clear whether MORFFO are a cause or a consequence of inversions. In the present study, we made a new finding that the positions of MORFFO were between palindrome sequences, especially near tRNA genes (Supplementary table 3). Transfer RNA genes are very important in structural variation in fern plastomes. For instance, inversion between *trnFM* and *trnE* is specific to moninophytes⁴⁸ and a number of inversions occurring near tRNA genes have been found in fern plastomes^{3,19,26,27}. In addition, length variations in noncoding regions of plastomes owing to tRNA gene repeats have been suggested^{19,27}. Transfer RNA-related genome rearrangements are not unique characteristic of

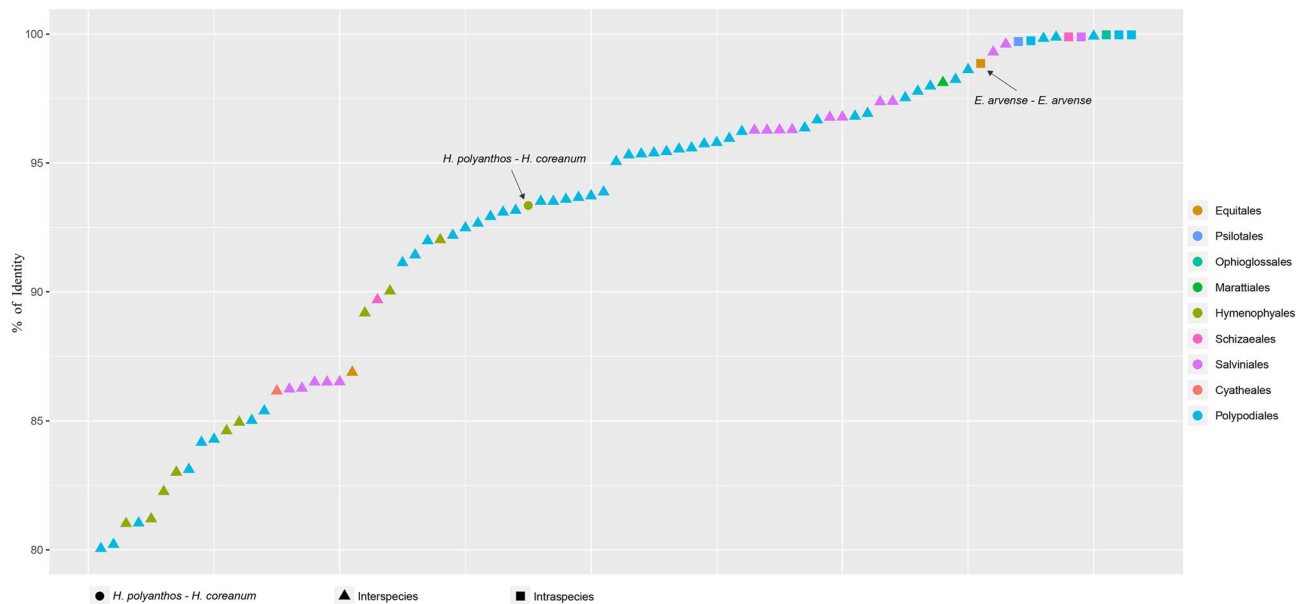


Figure 5. Percentage of sequence identity between two plastomes. Circle refers to the sequence identity between *H. polyanthos* and *H. coreanum*. Triangle and square refer to the sequence identity between two plastomes of interspecific level in the same genus and that of intraspecific level in the same species, respectively.

fern plastomes. Intermolecular recombination between distinct tRNA genes has been found in cereals⁴⁹, and extensive rearrangements were also suggested due to repeats or tRNA genes in *Trachelium caeruleum*⁵⁰. Consequently, tRNA genes in plastomes of ferns seem to be strongly associated with genome instability events which can lead to insertion of MORFFO using double-strand breaks⁵¹.

In bacterial genome evolution, the most transposable elements have short terminal inverted repeats⁵². The insertion of some specific sequences related to repetitive extragenic palindromic (REP) elements have been reported^{53,54}. Therefore, the mobility of MORFFO within the plastome is likely to be associated to the palindromic sequences at both ends of insertion regions. Two MORFFO in distant positions in a plastome of certain species in Polypodiales⁴⁷ and Hymenophyllales in this study also seem to be caused by the palindrome sequences.

Correctively, it is likely that the mobile elements, consisting of palindrome sequences, lead to the insertion of MORFFO in the plastomes of ferns, and genome instability caused by hairpin structures at the ends of MORFFO influence genome rearrangements on the plastome. As a result, inversions seem to frequently occur in the loci having MORFFO.

Degradation after insertion of long DNA contig including MORFFO. So far, more than one hundred plastome sequences in ferns have been reported, and MORFFO have been found within at least one species in Marattiales, Ophioglossales, Hymenophyllales, Gleicheniales, Schizaeales, Cyatheales, and Polypodiales (Supplementary Table 2). The MORFFO status (present or absent) varies at the family level in Polypodiales⁴⁷ or at the genus level in Ophioglossaceae, Schizaeaceae, Blechnaceae, Dryopteridaceae, Lomariopsidaceae, Polypodiaceae, Pteridaceae, and Thelypteridaceae. In addition, the inserted region of MORFFO also differs in certain genera²⁵ (i.e., *Asplenium*, *Cheilanthes*, *Myriopteris*). It is uncertain exactly when MORFFO were inserted in the fern plastomes. However, if MORFFO were independent events at different lineages of ferns, it is difficult to explain (1) why expanded regions, including MORFFO, have high sequence similarities throughout the fern lineages²⁵ and (2) how MORFFO were frequently transferred into the highly conserved genome, in terms of gene transfer⁵⁵. In addition, MORFFO-like sequences found in cyanobacteria or green algae imply that MORFFO come from a single origin that was started from nuclear genomes after endosymbiotic or horizontal gene transfer^{3,25}.

Even though MORFFO are supposed to be under selective pressure²⁵, they are not assumed to play a vital role for fern plastomes because there was no vestige of MORFFO in various fern families (Supplementary Table 2). It is not surprising to degrade unnecessary genes in a plastome of land plants^{56,57}. Especially, the degradation of *ndh* genes in Orchidaceae showed how easy it is to delete genes in plastomes among closely related taxa⁵⁸. MORFFO in the plastome of *Mankyua* and that in the mitochondrial genome of *Helminthostachys*³ also suggest MORFFO degradation through intracellular gene transfer.

In Hymenophyllaceae, the length of *rrn16-trnV* of *H. wrightii* is slightly expanded similar to previously reported plastomes^{20,59}, except *Vandenboschia*⁶⁰, owing to MORFFO. Interestingly, *rrn16-trnV* is almost tripled in *H. barbatum* and *H. coreanum* or quadrupled in *H. polyanthos* compared with *H. wrightii* (Fig. 2). As mentioned earlier, MORFFO seems to have existed in the ancestor of *Hymenophyllum*, the length variation of *rrn16-trnV* in *Hymenophyllum* is able to be interpreted into a consequence of degradation.

Similar to the plastomes in Ophioglossaceae and *Hymenophyllum*, MORFFO survived or disappeared during the evolution of fern plastomes. Partial sequences of MORFFO in most core leptosporangiate ferns except Pteridaceae²⁵ are consistent with the degradation hypothesis. MORFFO-lacking lineages may come from an

ancestor that has undergone an independent MORFFO degradation event. However, it is not clear why MORFFO remain in relatively recently diverged families in the fern phylogeny¹⁷, Pteridaceae of Polypodiales²⁵, if they do not have any function. One hypothesis is that *morffo1*, *morffo2*, and *morffo3*, which are found in Pteridaceae, are not intact genes but partial genes of a large ORF. Robison et al.²⁵ described that these three ORFs cluster immediately adjacent to one another when they are present, and *morffo1* and *morffo2* often form a larger ORF. ORF_B1 in *H. barbatum* is highly similar to *morffo1* and *morffo2* consistent with their observation. If insertion of MORFFO contig was the consequence of genomic instability by stem-loop structures and then rearrangements including insertion/deletion and inversion were caused by genomic instability at the inserted region, a larger ORF can be fragmented and shuffling of smaller fragments can make novel chimeric ORFs. Actually, these novel chimeric ORFs have been reported in the mitochondrial genomes of land plants^{61,62}.

Plastomes of eusporangiate and basal leptosporangiate ferns will shed light on the prototype of extant MORFFO, and this information will lead us to understand the evolution of MORFFO in fern plastomes.

Materials and Methods

Plant materials and DNA extraction. Four species, *H. barbatum* (CBNU-2018-0021), *H. polyanthos* (CBNU-2018-0187), *H. wrightii* (CBNU-2018-0330), and *H. coreanum* (CBNU-2018-0190), were sampled at Cheju Island and Jirisan in Korea. All specimens and living materials were deposited at the herbarium and greenhouse of Chungbuk National University. Using a fresh leaf, genomic DNAs were extracted by means of DNeasy Plant Mini Kit (Qiagen, Inc.) following the protocol proposed by the manufacturer.

Sequencing and assembling. Four high-quality DNAs were sequenced by Illumina HSeq X Ten. The raw reads were trimmed by trimmomatic 0.36⁶³ with the options of leading:10, trailing:10, slidingwindow:4:20, and min-len:50. The assembly method was identical to Kim et al.⁶⁴ using *Hymenophyllum hollochilum* (NC_039753) as a reference sequence.

Gene annotation. Genes in four plastome sequences were annotated compared with the genes of *H. hollochilum* using Geneious 10.2.6. Transfer RNA genes were verified using tRNAscan-SE⁶⁵ and ORFs with > 303 bp in length were also annotated using Geneious 10.2.6.

Analyses of ORFs in *Hymenophyllum* and expanded noncoding regions in fern plastomes. To investigate the ORFs located in the *rrn16-trnV* IGS having high length variation in the genus, we extracted that the region from four plastome sequences. First, *rrn16-trnV* IGS was compared to other nucleotides using blastn⁶⁶ to find out the homologous sequences. Thereafter, ORFs were searched using blastp to confirm the origin of these ORFs. To explore the physical positions of MORFFO, 126 plastome sequences were downloaded from NCBI and eight plastome sequences were received from Lehtonen and Cárdenas⁴⁷ (Supplementary Table 2).

Phylogenetic relationships among four *Hymenophyllum* species. A total of 85 coding genes, except duplicated genes, were extracted from eight Hymenophyllaceae species and two Osmundales and one Gleicheniales species. To obtain the best partitioning scheme for each gene position in a single concatenated sequence, ModelFinder⁶⁷ and PartitionFinder⁶⁸ were used for ML analysis and BI analysis, respectively. MP and ML analyses were conducted using PAUP⁶⁹ and IQ-TREE⁷⁰ with 1000 bootstrap replications. BI analysis was inferred by MrBayes⁷¹ under selected models in PartitionFinder with ngen = 1,000,000, samplefreq = 500, burninfrac = 0.25.

Estimation of DNA amounts. The genome size of *H. polyanthos* and *H. coreanum* was measured using living materials. For *H. polyanthos*, three populations were collected from Mt. Jiri-san in mainland of Korean peninsula and Mt. Halla-san of Jeju Island located at the most southern region of Korea. Two populations of *H. coreanum* were also collected from Mt. Jiri-san and Mt. Gariwang-san of southern and northern part of Korea, respectively. For a calibration standard, we grew *Nicotiana tabacum* in a greenhouse for weeks and used only 1 × 1 cm² of young leaf. The reference and two samples were separately chopped with 500 µl of Nuclei extraction buffer in CyStain UV Precise P kit (Sysmex Partec, Germany) in a Petri dish on ice. The extraction time for samples was extended for eight minutes because the numbers of events less than that time were very low to visualize the graph. The debris was filtered using non-sterile CellTrics filter Green 30 µm (Sysmex Partec, Germany) and nuclei were stained by Staining Buffer in CyStain UV Precise P kit (Sysmex Partec, Germany). Particles of each sample were measured using CyFlow Cube 6 flow cytometer (Sysmex Partec, Germany).

Data availability

The complete sequence data used in the current study are available in the NCBI GenBank repository. All data generated or analyzed during this study are included in this published article and its Supplementary Information files.

Received: 16 October 2019; Accepted: 12 June 2020

Published online: 06 July 2020

References

- Ruhlman, T. A. & Jansen, R. K. The plastid genomes of flowering plants. *Methods Mol. Biol.* **1132**, 3–38. https://doi.org/10.1007/978-1-62703-995-6_1 (2014).
- Viruel, J. et al. Phylogenetics, classification and typification of extant horsetails (Equisetum, Equisetaceae). *Bot J Linn Soc.* **189**, 311–352. <https://doi.org/10.1093/botlinnean/boz002> (2019).

3. Kim, H. T. & Kim, K. J. Evolution of six novel ORFs in the plastome of Mankyua chejuense and phylogeny of eusporangiate ferns. *Sci. Rep.* **8**, 16466. <https://doi.org/10.1038/s41598-018-34825-6> (2018).
4. Smith, D. R. Unparalleled GC content in the plastid DNA of Selaginella. *Plant Mol. Biol.* **71**, 627–639. <https://doi.org/10.1007/s11103-009-9545-3> (2009).
5. Wolf, P. G. *et al.* The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). *Gene* **350**, 117–128. <https://doi.org/10.1016/j.gene.2005.01.018> (2005).
6. Funk, H. T., Berg, S., Krupinska, K., Maier, U. G. & Krause, K. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol.* **7**, 45. <https://doi.org/10.1186/1471-2229-7-45> (2007).
7. McNeal, J. R., Kuehl, J. V., Boore, J. L. & de Pamphilis, C. W. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biol.* **7**, 57. <https://doi.org/10.1186/1471-2229-7-57> (2007).
8. Guisinger, M. M., Kuehl, J. V., Boore, J. L. & Jansen, R. K. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. *Mol. Biol. Evol.* **28**, 583–600. <https://doi.org/10.1093/molbev/msq229> (2011).
9. Cai, Z. *et al.* Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *J. Mol. Evol.* **67**, 696–704. <https://doi.org/10.1007/s00239-008-9180-7> (2008).
10. Knox, E. B. The dynamic history of plastid genomes in the Campanulaceae sensu lato is unique among angiosperms. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11097–11102. <https://doi.org/10.1073/pnas.1403363111> (2014).
11. Young, H. A., Lanzatella, C. L., Sarath, G. & Tobias, C. M. Chloroplast genome variation in upland and lowland switchgrass. *PLoS ONE* **6**, e23980. <https://doi.org/10.1371/journal.pone.0023980> (2011).
12. Mariotti, R., Cultrera, N. G., Diez, C. M., Baldoni, L. & Rubini, A. Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. *BMC Plant Biol.* **10**, 211. <https://doi.org/10.1186/1471-2229-10-211> (2010).
13. Tang, J. *et al.* A comparison of rice chloroplast genomes. *Plant Physiol.* **135**, 412–420. <https://doi.org/10.1104/pp.103.031245> (2004).
14. Doorduyn, L. *et al.* The complete chloroplast genome of 17 individuals of pest species *Jacobaea vulgaris*: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. *DNA Res.* **18**, 93–105. <https://doi.org/10.1093/dnares/dsr002> (2011).
15. Kim, H. T. & Kim, K. J. Chloroplast genome differences between Asian and American Equisetum arvense (Equisetaceae) and the origin of the hypervariable trnY-trnE intergenic spacer. *PLoS ONE* **9**, e103898. <https://doi.org/10.1371/journal.pone.0103898> (2014).
16. Wolf, P. G., Rowe, C. A., Sinclair, R. B. & Hasebe, M. Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern *Adiantum capillus-veneris* L. *DNA Res.* **10**, 59–65. <https://doi.org/10.1093/dnares/10.2.59> (2003).
17. Smith, A. R. *et al.* A classification for extant ferns. *Taxon* **55**, 705–731. <https://doi.org/10.2307/25065646> (2006).
18. Grewe, F., Guo, W., Gubbels, E. A., Hansen, A. K. & Mower, J. P. Complete plastid genomes from *Ophioglossum californicum*, *Psilotum nudum*, and *Equisetum hyemale* reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. *BMC Evol. Biol.* **13**, 8. <https://doi.org/10.1186/1471-2148-13-8> (2013).
19. Kim, H. T., Chung, M. G. & Kim, K. J. Chloroplast genome evolution in early diverged leptosporangiate ferns. *Mol. Cells* **37**, 372–382. <https://doi.org/10.14348/molcells.2014.2296> (2014).
20. Kuo, L. Y., Qi, X., Ma, H. & Li, F. W. Order-level fern plastome phylogenomics: new insights from Hymenophyllales. *Am. J. Bot.* **105**, 1545–1555. <https://doi.org/10.1002/ajb2.1152> (2018).
21. Gao, L. *et al.* Plastome sequences of *Lygodium japonicum* and *Marsilea crenata* reveal the genome organization transformation from basal ferns to core leptosporangiates. *Genome Biol. Evol.* **5**, 1403–1407. <https://doi.org/10.1093/gbe/evt099> (2013).
22. Zhong, B., Fong, R., Collins, L. J., McLenachan, P. A. & Penny, D. Two new fern chloroplasts and decelerated evolution linked to the long generation time in tree ferns. *Genome Biol. Evol.* **6**, 1166–1173. <https://doi.org/10.1093/gbe/evu087> (2014).
23. Lu, J.-M., Zhang, N., Du, X.-Y., Wen, J. & Li, D.-Z. Chloroplast phylogenomics resolves key relationships in ferns. *J. System. Evol.* **53**, 448–457. <https://doi.org/10.1111/jse.12180> (2015).
24. Wei, R. *et al.* Plastid phylogenomics resolve deep relationships among Eupolypod II Ferns with rapid radiation and rate heterogeneity. *Genome Biol. Evol.* **9**, 1646–1657. <https://doi.org/10.1093/gbe/evx107> (2017).
25. Robison, T. A. *et al.* Mobile elements shape plastome evolution in ferns. *Genome Biol. Evol.* **10**, 2558–2571. <https://doi.org/10.1093/gbe/evy189> (2018).
26. Wolf, P. G., Roper, J. M. & Duffy, A. M. The evolution of chloroplast genome structure in ferns. *Genome* **53**, 731–738. <https://doi.org/10.1139/g10-061> (2010).
27. Gao, L., Zhou, Y., Wang, Z. W., Su, Y. J. & Wang, T. Evolution of the rpoB-psbZ region in fern plastid genomes: notable structural rearrangements and highly variable intergenic spacers. *BMC Plant Biol.* **11**, 64. <https://doi.org/10.1186/1471-2229-11-64> (2011).
28. Pryer, K. M., Smith, A. R., Hunt, J. S. & Dubuisson, J. Y. rbcL data reveal two monophyletic groups of filmy ferns (Filicopsida: Hymenophyllaceae). *Am. J. Bot.* **88**, 1118–1130 (2001).
29. Ebihara, A., Dubuisson, J.-Y., Iwatsuki, K., Hennequin, S. & Ito, M. A taxonomic revision of Hymenophyllaceae. *Blumea-Biodivers Evol Biogeogr Plants* **51**, 221–280 (2006).
30. Morton, C. V. The Genera, Subgenera, and sections of the Hymenophyllaceae. *Contrib. U.S. Natl. Herb.* **38**, 153–214 (1968).
31. Iwatsuki, K. The Hymenophyllaceae of Asia, excluding Malaysia. *J. Fac. Sci. U. Tokyo* **III**(13), 501–551 (1985).
32. Copeland, E. B. *Genera Filicum: the genera of ferns*. (Chronica Botanica Co., Waltham, 1947).
33. Copeland, E. B. Genera Hymenophyllacearum. *Philos. J. Sci.* **67**, 1–110 (1938).
34. contribution of morphology and cytology. Hennequin, S. Phylogenetic relationships within the fern genus *Hymenophyllum* s.l. (Hymenophyllaceae, Filicopsida). *C.R. Biol.* **326**, 599–611. [https://doi.org/10.1016/s1631-0691\(03\)00146-x](https://doi.org/10.1016/s1631-0691(03)00146-x) (2003).
35. Iwatsuki, K. Studies in the systematics of filmy ferns: VII. A scheme of classification based chiefly on the Asiatic species. *Acta Phytotaxonomica et Geobotanica* **35**, 165–179 (1984).
36. Hennequin, S., Ebihara, A., Ito, M., Iwatsuki, K. & Dubuisson, J.-Y. New insights into the phylogeny of the genus *Hymenophyllum* sl (Hymenophyllaceae): revealing the polyphyly of *Mecodium*. *Syst. Bot.* **31**, 271–284 (2006).
37. Nakai, T. Notes on Japanese Ferns III. *Shokubutsugaku Zasshi* **40**, 239–275 (1926).
38. Ebihara, A. *The Standard of Ferns and Lycophytes in Japan* (Gakken Plus, Tokyo, 2016).
39. Ebihara, A., Nitta, J. H. & Iwatsuki, K. The Hymenophyllaceae of the Pacific area. 2. *Hymenophyllum* (excluding subgen. *Hymenophyllum*). *Bull. Natl. Museum Nat. Sci. Ser. B* **36**, 43–59 (2010).
40. Vasques, D. T., Ebihara, A., Hirai, R. Y., Prado, J. & Motomi, I. Phylogeny of *Hymenophyllum* subgen. *Mecodium* (Hymenophyllaceae), with special focus on the diversity of the *Hymenophyllum* polyanthos species complex. *Plant System. Evol.* <https://doi.org/10.1007/s00606-019-01609-y> (2019).
41. Hoang, N. V., Furtado, A., McQualter, R. B. & Henry, R. J. Next generation sequencing of total DNA from sugarcane provides no evidence for chloroplast heteroplasmy. *New Negatives Plant Sci.* **1**, 33–45 (2015).
42. Leitch, I. J. *et al.* The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Ann. Bot.* **101**, 805–814. <https://doi.org/10.1093/aob/mcm326> (2008).
43. Clark, J. *et al.* Genome evolution of ferns: evidence for relative stasis of genome size across the fern phylogeny. *New Phytol.* **210**, 1072–1082. <https://doi.org/10.1111/nph.13833> (2016).

44. Hennequin, S., Ebihara, A., Dubuisson, J.-Y. & Schneider, H. Chromosome number evolution in Hymenophyllum (Hymenophyllaceae), with special reference to the subgenus Hymenophyllum. *Mol. Phylogenet. Evol.* **55**, 47–59. <https://doi.org/10.1016/j.ympev.2010.01.001> (2010).
45. Brownsey, P. & Perrie, L. Hymenophyllaceae. In *Flora of New Zealand—Ferns and Lycophytes. Fascicle 15* (eds Breitwieser, I. et al.) (Manaaki Whenua Press, Lincoln, 2016).
46. Iwatsuki, K. *Pteridophytes and Gymnosperms* 157–163 (Springer, Berlin, 1990).
47. Lehtonen, S. & Cárdenas, G. G. Dynamism in plastome structure observed across the phylogenetic tree of ferns. *Bot J Linn Soc.* **190**, 229–241 (2019).
48. Karol, K. G. et al. Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. *BMC Evol. Biol.* **10**, 321. <https://doi.org/10.1186/1471-2148-10-321> (2010).
49. Hiratsuka, J. et al. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet. MGG* **217**, 185–194 (1989).
50. Haberle, R. C., Fourcade, H. M., Boore, J. L. & Jansen, R. K. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J. Mol. Evol.* **66**, 350–361. <https://doi.org/10.1007/s00239-008-9086-4> (2008).
51. Aguilera, A. & Gómez-González, B. Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Genet.* **9**, 204 (2008).
52. Darmon, E. & Leach, D. R. F. Bacterial genome instability. *Microbiol. Mol. Biol. Rev.* **78**, 1–39. <https://doi.org/10.1128/mbr.00035-13> (2014).
53. Clément, J.-M., Wilde, C., Bachellier, S., Lambert, P. & Hofnung, M. IS1397 is active for transposition into the chromosome of *Escherichia coli* K-12 and inserts specifically into palindromic units of bacterial interspersed mosaic elements. *J. Bacteriol.* **181**, 6929–6936 (1999).
54. Tobes, R. & Pareja, E. Bacterial repetitive extragenic palindromic sequences are DNA targets for insertion sequence elements. *BMC Genom.* **7**, 62. <https://doi.org/10.1186/1471-2164-7-62> (2006).
55. Leister, D. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends Genet.* **21**, 655–663. <https://doi.org/10.1016/j.tig.2005.09.004> (2005).
56. Logacheva, M. D., Schelkunov, M. I., Nuraliev, M. S., Samigullin, T. H. & Penin, A. A. The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. *Genome Biol. Evol.* **6**, 238–246. <https://doi.org/10.1093/gbe/evu001> (2014).
57. Revill, M. J., Stanley, S. & Hibberd, J. M. Plastid genome structure and loss of photosynthetic ability in the parasitic genus *Cuscuta*. *J. Exp. Bot.* **56**, 2477–2486. <https://doi.org/10.1093/jxb/eri240> (2005).
58. Kim, H. T. & Chase, M. W. Independent degradation in genes of the plastid *ndh* gene family in species of the orchid genus *Cymbidium* (Orchidaceae; Epidendroideae). *PLoS ONE* **12**, e0187318. <https://doi.org/10.1371/journal.pone.0187318> (2017).
59. Lehtonen, S. The complete plastid genome sequence of *Trichomanes trollii* (Hymenophyllaceae). *Nordic J. Bot.* **36**, e02072. <https://doi.org/10.1111/njb.02072> (2018).
60. Ruiz-Ruano, F. J., Navarro-Domínguez, B., Camacho, J. P. M. & Garrido-Ramos, M. A. Full plastome sequence of the fern *Vandenboschia speciosa* (Hymenophyllales): structural singularities and evolutionary insights. *J. Plant. Res.* **132**, 3–17. <https://doi.org/10.1007/s10265-018-1077-y> (2019).
61. Tanaka, Y., Tsuda, M., Yasumoto, K., Yamagishi, H. & Terachi, T. A complete mitochondrial genome sequence of Ogura-type male-sterile cytoplasm and its comparative analysis with that of normal cytoplasm in radish (*Raphanus sativus* L.). *BMC Genom.* **13**, 352. <https://doi.org/10.1186/1471-2164-13-352> (2012).
62. Jo, Y. D., Choi, Y., Kim, D. H., Kim, B. D. & Kang, B. C. Extensive structural variations between mitochondrial genomes of CMS and normal peppers (*Capsicum annuum* L.) revealed by complete nucleotide sequencing. *BMC Genom.* **15**, 561. <https://doi.org/10.1186/1471-2164-15-561> (2014).
63. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> (2014).
64. Kim, H. T. et al. Seven New Complete Plastome Sequences Reveal Rampant Independent Loss of the *ndh* Gene Family across Orchids and Associated Instability of the Inverted Repeat/Small Single-Copy Region Boundaries. *PLoS ONE* **10**, e0142215. <https://doi.org/10.1371/journal.pone.0142215> (2015).
65. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucl. Acids Res.* **25**, 955–964 (1997).
66. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) (1990).
67. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587 (2017).
68. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773 (2016).
69. Swofford, D. *PAUP 4.0: phylogenetic analysis using parsimony* (Smithsonian Institution, Washington, 1998).
70. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2014).
71. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542. <https://doi.org/10.1093/sysbio/sys029> (2012).

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) under grant no. 2018R1D1A3B07048213. We really appreciate Sang Hee Park and Kanghyup Lee for helping in the sampling of *Hymenophyllum* species. We also thank Tanner A. Robison and Glenda G. Cárdenas for giving information about fern plastome analysis.

Author contributions

H.T.K. performed research design, writing manuscript, labwork, and data analysis. J.S.K. contributed to research design and writing manuscript.

Competing interests

The authors declare no competing interests

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-68000-7>.

Correspondence and requests for materials should be addressed to J.S.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020