



OPEN

Complete organelle genomes of Korean fir, *Abies koreana* and phylogenomics of the gymnosperm genus *Abies* using nuclear and cytoplasmic DNA sequence data

Seongjun Park¹, Myounghai Kwak^{2✉} & SeonJoo Park^{3✉}

Abies koreana E.H.Wilson is an endangered evergreen coniferous tree that is native to high altitudes in South Korea and susceptible to the effects of climate change. Hybridization and reticulate evolution have been reported in the genus; therefore, multigene datasets from nuclear and cytoplasmic genomes are needed to better understand its evolutionary history. Using the Illumina NovaSeq 6000 and Oxford Nanopore Technologies (ONT) PromethION platforms, we generated complete mitochondrial (1,174,803 bp) and plastid (121,341 bp) genomes from *A. koreana*. The mitochondrial genome is highly dynamic, transitioning from *cis*- to *trans*-splicing and breaking conserved gene clusters. In the plastome, the ONT reads revealed two structural conformations of *A. koreana*. The short inverted repeats (1186 bp) of the *A. koreana* plastome are associated with different structural types. Transcriptomic sequencing revealed 1356 sites of C-to-U RNA editing in the 41 mitochondrial genes. Using *A. koreana* as a reference, we additionally produced nuclear and organelle genomic sequences from eight *Abies* species and generated multiple datasets for maximum likelihood and network analyses. Three sections (*Balsamea*, *Momi*, and *Pseudopicea*) were well grouped in the nuclear phylogeny, but the phylogenomic relationships showed conflicting signals in the mitochondrial and plastid genomes, indicating a complicated evolutionary history that may have included introgressive hybridization. The obtained data illustrate that phylogenomic analyses based on sequences from differently inherited organelle genomes have resulted in conflicting trees. Organelle capture, organelle genome recombination, and incomplete lineage sorting in an ancestral heteroplasmic individual can contribute to phylogenomic discordance. We provide strong support for the relationships within *Abies* and new insights into the phylogenomic complexity of this genus.

Abbreviations

ETS	External transcribed spacer
ITS	Internal transcribed spacer
MINCs	Mitochondrial DNA of nuclear origin
MIPTs	Mitochondrial DNA of plastid origin
ORF	Open reading frame
LSC	Large single copy
SSC	Small single copy
IR	Inverted repeat

¹Institute of Natural Science, Yeungnam University, Gyeongsan, Gyeongbuk 38541, South Korea. ²National Institute of Biological Resources, Incheon 22689, South Korea. ³Department of Life Sciences, Yeungnam University, Gyeongsan, Gyeongbuk 38541, South Korea. ✉email: mhkwak1@korea.kr; sjpark01@ynu.ac.kr

Abies Mill. (Fir) is a genus of 48 evergreen conifers in the family Pinaceae that is native to North and Central America, Europe, Asia, and North Africa¹. The genus *Abies* is distinguished from the other ten Pinaceae genera by morphological traits, including erect seed cones and deciduous seed scales at maturity with a persistent cone rachis¹. Based on analyses of the internal transcribed spacer (ITS) region of nuclear ribosomal DNA (nrDNA), single-copy nuclear genes, plastid DNA, and mitochondrial DNA sequences^{2–6}, *Abies* is a monophyletic group sister to the genus *Keteleeria*. The monophyly of sections is well supported, but the evolutionary relationships within the sections are not fully resolved due to low bootstrap values or polytomies. Recurrent hybridization influences the critical evolutionary processes within the genus *Abies*^{2,7,8}. Genomic resources can provide the opportunity for a comprehensive understanding of the origin and evolution of this hierarchical complexity. Recent research has examined the genome-scale biogeography of the genus *Abies*; however, only the mitogenome has been used⁹.

Gymnosperms comprise 12 families and 83 genera containing approximately 1079 species of Cycadophyta (cycads), Ginkgophyta, Pinophyta, and Gnetophyta¹⁰. These groups show dynamic modes of inheritance in organelle genomes. For example, both organelle genomes of cycads, *Ginkgo*, and gnetophytes, are maternally inherited^{11,12}. However, most species of Pinaceae and Taxaceae (Pinophyta) show maternal inheritance of the mitogenome but paternal inheritance of the plastome¹¹. Because of this, it is challenging to infer the evolutionary history of Pinaceae and Taxaceae using only one genome. Therefore, three genomic sequences are required to pinpoint maternal, paternal, and biparental inheritance patterns to fully understand their evolutionary histories.

Next-generation sequencing (NGS) techniques and bioinformatic tools provide genomic-scale data for genomic and phylogenomic studies at a wide range of taxonomic levels. At least 200 gymnosperm plastomes have been sequenced, but only a few mitogenomes are available yet according to the NCBI Genome database, which was accessed on December 11, 2023. These genomic data show that gymnosperm plastomes and mitogenomes are diverse in their structural evolution. The plastomes of cycads, *Ginkgo*, and gnetophytes contain a typical inverted region (IR), whereas those of Pinophyta lacks a large IR^{13–15}. However, the plastomes of Pinophyta have highly reduced repeats that generate substoichiometric isomers by homologous recombination¹³. Gymnosperm plastomes contain 66–87 protein-coding genes, 28–35 tRNA genes, and four rRNA genes, ranging from 85.3 kb (*Parasitaxus* in Pinophyta) to 166.3 kb (*Macrozamia* in cycads) in size, and exhibit inversion, relocation, and losses of genes^{16–18}. In contrast, only five gymnosperm mitogenomes have been assembled into circular structures (*Cycas taitungensis*¹⁹, *Cycas debaoensis*²⁰, *Ginkgo biloba*²¹, *Welwitschia mirabilis*²¹, and *Taxus cuspidate*²²), and eight have been assembled into linear scaffolds or both linear and circular structures (three *Abies* [see below], *Larix sibirica*²³, *Picea abies*²⁴, *Picea sitchensis*²⁵, *Picea glauca*²⁶, and *Pinus taeda*²⁷). The sizes of the mitogenomes are also variable, ranging from 346.5 kb in *G. biloba*²¹ to 11.69 Mb in *L. sibirica*²³ from Pinophyta and contain 29–41 protein-coding genes, 28–35 tRNA genes, and three rRNA genes. They show various features, including mitochondrial DNA of plastid (MIPTs) and nuclear (MINCs) origin, repeats, and RNA editing^{19,22,28,29}.

To date, 43 plastomes from 25 *Abies* taxa have been sequenced based on the NCBI Genome database, which was accessed on December 11, 2023, ranging from 119.4 kb (*A. religiosa*) to 121.8 kb (*A. fargessii*) and containing 72 protein-coding genes with 16 introns. All 11 *ndh* genes have been lost in these plastomes. However, draft mitogenome sequences of only three species of *Abies* have been reported: *A. alba* assembled into 11 scaffolds (1.43 Mb)³⁰, *A. firma* assembled into 172 scaffolds (1.33 Mb)²⁸, and *A. sibirica* assembled into 237 scaffolds (1.49 Mb)³¹. These genomic data suggest that the ancestral *Abies* mitogenome contains 41 protein genes with 26 introns.

PacBio SMRT or Oxford Nanopore Technologies (ONT) sequencing provides opportunities to assemble large and complex genomes with long reads²⁵, but it is still challenging to complete. To improve our understanding of the evolution of organelle genomes in the genus *Abies*, we sequenced the complete mitochondrial and plastid genomes from Korean fir, *Abies koreana* and characterized MIPTs and repeats in the complete mitogenome and RNA editing in 41 mitochondrial genes. To reconstruct a genome-scale phylogeny for the *Abies*, we sequenced complete plastomes and extracted these mitochondrial genes from Illumina assemblies of eight additional species. In addition, we analyzed the nuclear single-copy genes as well as ribosomal DNA (rDNA) clusters, which include loci for the 5' external transcribed spacer (5'-ETS), 18S, ITS1, 5.8S, ITS2, and 26S rDNA regions. Finally, we used these data to explore the phylogenomic evidence for the selected *Abies* relationships.

Results

Mitochondrial and plastid genome organization of *A. koreana*

Using Illumina NovaSeq 6000 and ONT PromethION sequencing, we obtained a total of 551,969,293 paired-end reads (100 bp × 2) and 1,865,257 long reads (average length 14,207 bp, range 1–3,387,768 bp) for *A. koreana*. The mitochondrial and plastid genomes of *A. koreana* were assembled by Velvet, Spades, and MaSuRCA using Illumina and Oxford reads with deep coverage (PE/ONT; 418/224-fold for mitogenome, PE/ONT; 1950/543-fold for plastome).

The *A. koreana* mitogenome was assembled into a circular molecule with a genome size of 1,174,803 bp (Table 1 and Fig. 1). The GC content was 45.9%, and the genome consisted of 3.08% protein-coding genes, 96.39% noncoding regions, 0.06% tRNA genes, and 0.47% rRNA genes (Table 1). The *A. koreana* mitogenome contained 153,910 bp of repetitive DNA (13.1%) (Table and Fig. 2a), ranging from 30 to 27,029 bp in length, including nine large (> 1000 bp), 36 intermediate (100–1000 bp), and 525 small (< 100 bp) repeats (Table S1). MIPT sequences were found in six fragments (504 bp) throughout the *A. koreana* mitogenome ranging from 68 to 106 bp in length (Table S2). Four partial genes (*petA*, *psaA*, *psbK*, and *ycf1*) and one tRNA gene (*trnW-CCA*) were identified. No *Bpu*-like elements were detected in the *A. koreana* mitogenome. The *A. koreana* mitogenome also contained 48,022 bp of diverse transposable elements (TEs) (Table S3). The majority of these TEs were long terminal repeat (LTR) retrotransposons (*copia*- and *gypsy*-like) (29,809 bp; 62.1%). The mitogenome of

	Mitochondrial genome	Plastid genome
Genome size (bp)	1,174,803	121,341
GC content (%)	45.9	38.3
Genes		
Protein-coding genes	41	72 (2)
(%)	3.08	50.29
tRNA genes	9	34 (1)
(%)	0.06	2.18
rRNA genes	3 (1)	4
(%)	0.47	3.73
Introns		
<i>cis</i> -spliced group	13	15
<i>trans</i> -spliced group	13	1
Plastid-derived		
Protein-coding genes	4	–
tRNA genes	1	–
Repeats (bp)	153,910	2357

Table 1. Characteristics of *Abies koreana* organelle genomes.

A. koreana contained 41 protein-coding genes, nine tRNAs, and three rRNAs (Table 1). In *Abies koreana*, 11 of the 14 clusters inferred from the ancestral angiosperm mitochondrial genome were missing: *nad3-rps12*, *rpl16-rps3-rps19-rpl2*, and *trnP-sdh3* (Table S4). Twenty-six group II introns were found in the ten genes, 13 and 13 of which were *cis*- and *trans*-spliced, respectively (Fig. 1). Six ORFs of at least 150 bp in length, two (*orf97* and *orf70*) of which were present as two copies, appeared to be chimeric ORFs containing small fragments (> 30 bp) of mitochondrial genes (Table S5). One (*orf69*) of these ORFs was predicted to encode one transmembrane helix containing small fragments of the *ccmC* gene (Fig. S1).

PREP-Mt with a cutoff of 0.2 predicted 1268 putative C-to-U RNA editing sites in the 41 *A. koreana* mitochondrial protein-coding genes (Tables S6 and S7). Mapping of the RNA-seq reads and alignment of transcripts revealed 1356 C-to-U RNA editing sites and confirmed the 1122 sites predicted by PREP-Mt for the 41 mitochondrial genes. Moreover, 116 silent RNA editing sites were detected (Table S7). The ten *A. koreana* mitochondrial genes (*atp6*, *cox1*, *mttB*, *nad1*, *nad6*, *nad9*, *rpl2*, *rps3*, *rps4*, and *rps10*) had an ACG start codon that was RNA edited to AUG in their transcript. In the *atp4*, *atp6*, *atp9*, *ccmFc*, *cox1*, *nad4L*, *nad4*, *rpl2*, *rpl16*, and *sdh3* genes, premature stop codons were created by C-to-U RNA editing. Among them, the *A. koreana* *cox1* gene showed 16 sites of C-to-U editing, creating a stop codon, and then 23 sites of C-to-U editing generated a new AUG start codon (Tables S6 and S7, and Fig. S2). The *A. koreana* *rpl16* gene had a GTG codon, and *rps19* had a GCG codon (RNA edited to GUG in its transcript) as an alternative start codon (Fig. S3). The gene most affected by RNA editing was *nad5*, for which 108 changes caused 98 codon changes (Table S6). The highest density of RNA editing sites was in *atp6*, in which 61 out of 777 nucleotides were edited (Table S6). Leucine was mostly affected by C-to-U editing, with 41.85% (519 sites) codon changes, and 23.95% (297 sites) of its amino acids were converted to phenylalanine (Table S6).

The plastome of *A. koreana* was 121,341 bp with a pair of inverted repeats of 1186 bp separated by a small single-copy (SSC) region (42,635 bp) and a large single-copy (LSC) region (76,334 bp) (Table 1 and Fig. 3a). Long reads provided strong evidence for the repeat-mediated rearrangement activity of the *A. koreana* plastome (Fig. 2b). Mapping of long reads to the plastome “form A” showed that many of the reads were consistent with the reference form except for the conflicting reads (Fig. 3b). The inconsistent areas were indicated with the corresponding regions of plastome “form B”, which identified the boundaries of the repeats (Fig. 3b). The *A. koreana* plastome contained 2357 bp of repetitive DNA (Table 1 and Fig. 3b), ranging from 32 to 1186 bp in length (Table S8). The GC content was 38.3%, and the genome consisted of 50.29% protein-coding genes, 43.80% noncoding regions, 2.18% tRNA genes, and 3.73% rRNA genes. The plastome contained 72 protein-coding genes, 34 tRNA genes, and four rRNA genes (Table 1). The short IR region contained the *trnS*-GCU, *psaM*, *ycf12*, and partial *trnG*-UCC genes (Fig. 3).

Abies plastid genomes, mitochondrial genes, and nuclear DNAs

We determined complete plastomes from eight additional species spanning three sections of *Abies*, including *Balsamea* (*A. sibirica*, *A. nephrolepis*, *A. sachalinensis*, and *A. veitchii*), *Momi* (*A. pindrow*, *A. firma*, and *A. kawakamii*), and *Pseudopicea* (*A. spectabilis*). The assembled *Abies* plastomes were similar in terms of their organization, size and gene/intron content to those of the *A. koreana* plastome (Table S9). In terms of gene content, the *A. kawakamii* plastome contained two copies of the *trnH*-GUG gene with 97.3% nucleotide sequence identity. The duplicated copy was located downstream from the *trnT*-GGU gene (approximately 31 kb from the original copy). Among the eight species of *Abies*, *A. kawakamii* had the largest (121,803 bp) plastome, and *A. spectabilis* had the smallest (119,914 bp) plastome. The short IR size ranged from 1172 (*A. spectabilis*) to 1186 bp (*A. firma*, *A. nephrolepis*, *A. sachalinensis*, and *A. veitchii*). Using the mitochondrial gene sequences

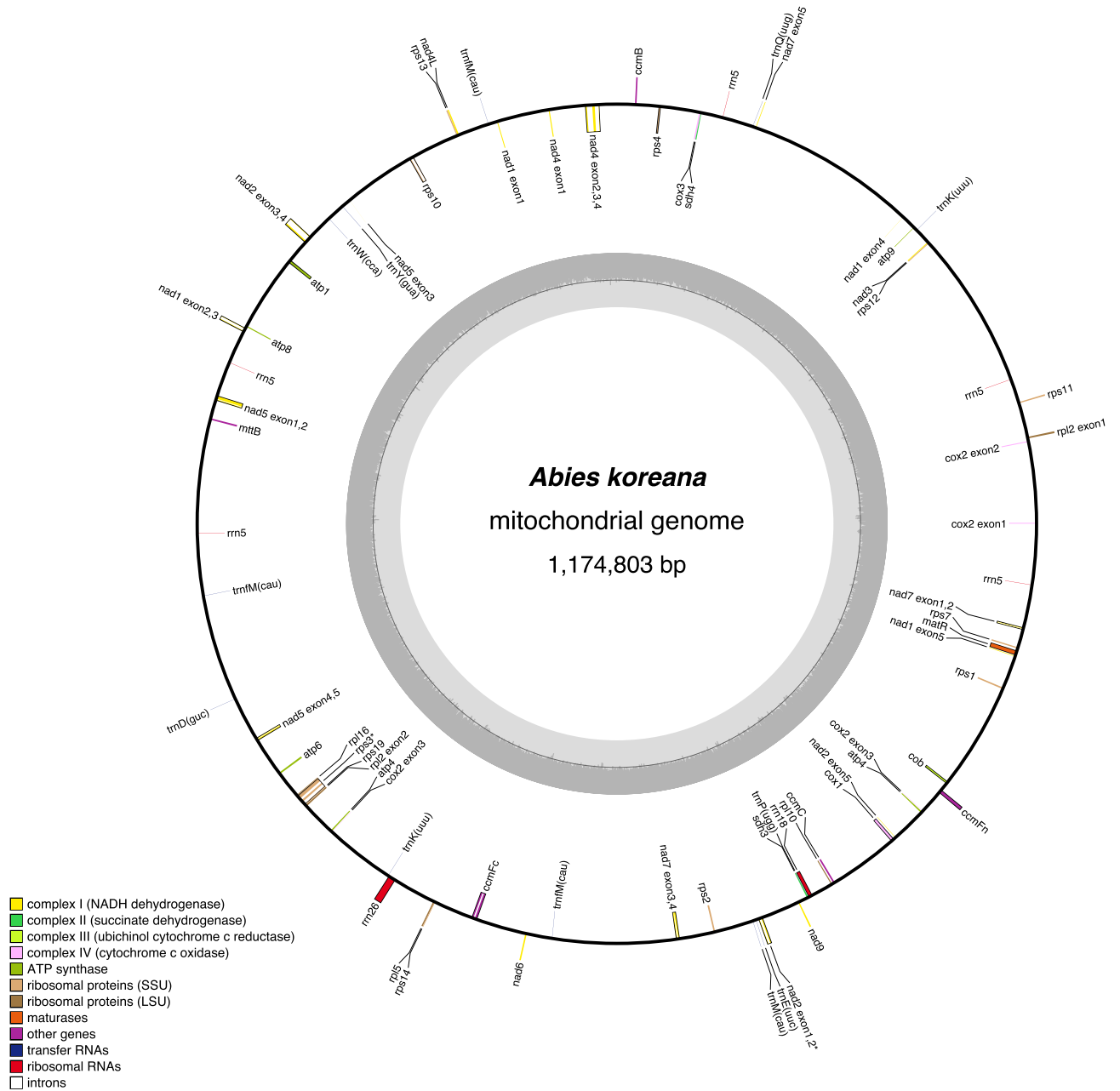


Figure 1. Circle map of the mitochondrial genome of *Abies koreana*. Genes on the inside and outside the map are transcribed in the clockwise and counterclockwise directions, respectively.

from *A. koreana*, we obtained 41 mitochondrial-encoded genes and 13 *cis*-spliced introns (*ccmFci829*, *nad1i477*, *nad2i156*, *nad2i709*, *nad4i1399*, *nad4i976*, *nad5i230*, *nad5i1872*, *nad7i140*, *nad7i676*, *rps3i74*, *rps3i257*, and *rps10i235*) for an additional eight *Abies* species.

We obtained nuclear single-copy gene sequences from the nine *Abies* PE reads using Read2Tree with a total of 1000 orthologs from *Picea glauca* and *Pinus taeda* and used only 452 single-copy orthologs that were present in the 11 analyzed species. In addition, we recovered nine nuclear rDNA sequences, including 5'-ETS, 18S, ITS1, 5.8S, ITS2, and 26S, which ranged in size from 7877 bp (*A. sibirica*) to 7906 bp (*A. kawakamii*) (Fig. S4). The length of ITS1 ranged from 1228 to 1257 bp, and that of ITS2 was 244 bp. The length of the partial 5'-ETS region ranged from 1043 to 1044 bp in length.

***Abies* phylogeny based on plastid, mitochondrial, and nuclear sequences**

To examine the phylogenetic relationships among the nine *Abies* species, we performed phylogenomic analyses using a concatenated alignment of plastid and mitochondrial protein-coding genes (plastid: 61,506 aligned nucleotide positions; mitochondrial: 34,501 aligned nucleotide positions), adding introns (plastid: 10,264 aligned nucleotide positions; mitochondrial: 23,302 aligned nucleotide positions), or a whole plastome alignment (122,389 aligned nucleotide positions) (Fig. 4a–g). The cladogram, based on plastid datasets, indicates that *Abies*

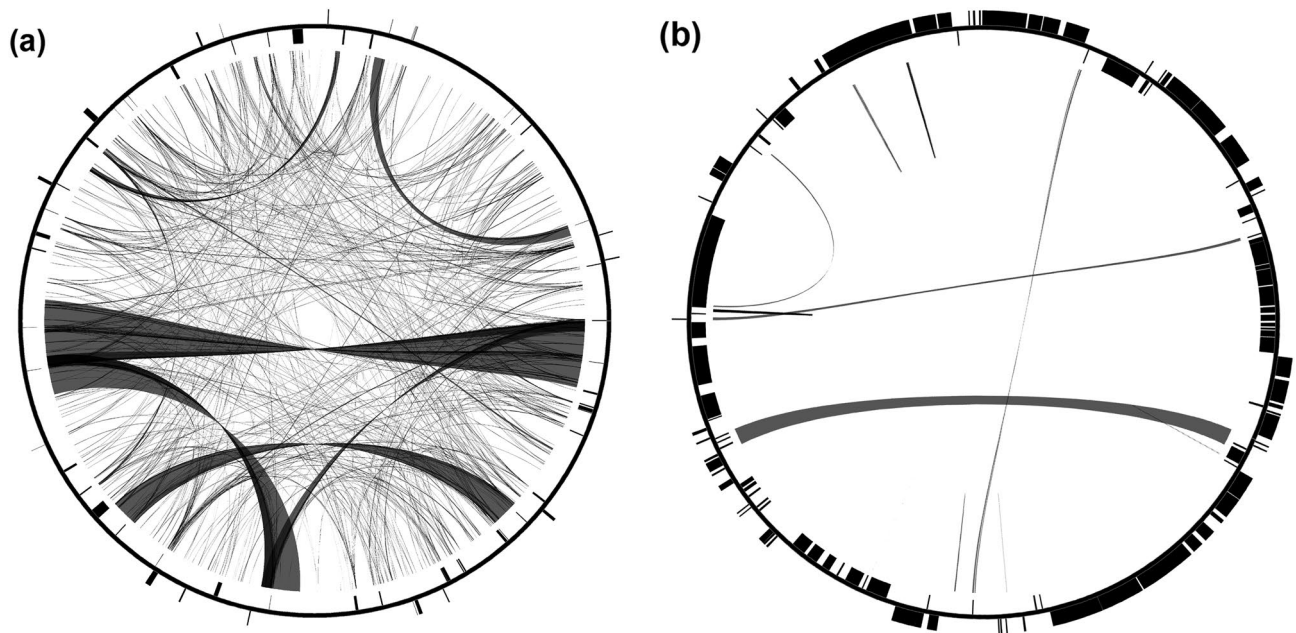


Figure 2. Distribution of repetitive DNA in *Abies koreana* organelle genomes. Within the circular maps, black lines represent the locations of pairs of repeats, with crossing lines denoting reverse repeats. In the inner and outer circles, the black boxes denote the locations of the mitochondrial genes. (a). mitochondrial genome. (b) plastid genome.

is strongly supported as monophyletic according to the bootstrap (BS) values of 100% and is sister to *Cedrus* (BS = 100%) (Fig. 4a–d) and contains two major clades. One clade comprising four species of sect. *Balsamea* except *A. sibirica* is strongly supported (BS = 100%). *Abies sibirica* is consistently nested within the other clade with strong support (BS = 100%, whole plastid dataset). However, there were conflicts in the relationships among *A. firma*, *A. spectabilis*, *A. pindrow*, *A. sibirica*, and *A. kawakamii* according to the datasets (Fig. 4a–d). The concatenated alignment of all 41 mitochondrial protein-coding genes and 11 introns revealed a topology with relatively strong bootstrap support (Fig. 4e–g). We excluded two mitochondrial introns (*ccmFci829* and *nad1i477*) due to the extreme size variation in the outgroups. However, we additionally generated a 41-gene/13-intron data matrix with 56,378 aligned nucleotide positions for only *Abies* species, showing that the topology is consistent with the phylogenetic tree, including outgroups (Fig. S5). In contrast to the plastid results, *A. sachalinensis* and *A. veitchii* of sect. *Balsamea* were sisters to *A. firma* of sect. *Momi* (BS = 99%). *Abies pindrow*, *A. spectabilis*, and *A. kawakamii* were sisters to *A. koreana*, *A. nephrolepis*, and *A. sibirica* (BS = 100%; Fig. S5). Moreover, *A. sibirica* was sister to *A. koreana* and *A. nephrolepis* with strong support (BS = 95%).

Phylogenomic analyses of the genus *Abies* based on nuclear rDNA sequences and single-copy gene sequences were also performed (Figs. 4h and 5). In particular, we conducted phylogenomic analyses of 452 nuclear single-copy genes employing coalescent- and concatenation-based methods. Our results revealed that the nuclear phylogenies were not congruent with those of the organellar topology. The cladogram, based on the nuclear rDNA datasets (7918 aligned nucleotide positions), provided strong support for the monophyly of *A. koreana*, *A. nephrolepis*, *A. sachalinensis*, *A. sibirica*, and *A. veitchii* from sect. *Balsamea* (BS = 100%) and for *A. firma*, *A. pindrow*, and *A. kawakamii* from sect. *Momi* (BS = 100%). A sister relationship between sect. *Momi* and *A. spectabilis* was strongly supported (BS = 100%). *Abies koreana* was sister to *A. sibirica* and *A. sachalinensis* (BS = 78%). The topologies of the Accurate Species TRee ALgorithm (ASTRAL) and the maximum likelihood (ML) supermatrix (441,999 aligned nucleotide positions) based on 452 genes were not identical (Fig. 5). The ASTRAL tree showed two major clades: one clade comprised five species of sect. *Balsamea* and *A. firma* from sect. *Momi* and the other was composed of *A. spectabilis*, *A. pindrow*, and *A. kawakamii*. In contrast to the ASTRAL tree, the concatenated ML tree showed that *A. spectabilis*, *A. pindrow*, and *A. kawakamii* were nested with the remaining *Abies* species.

To test phylogenetic incongruence, we compared four topologies (the whole plastome, 41 + 13 mitochondrial, nuclear rDNA and 452 nuclear gene sequences) by performing the approximately unbiased (AU), Shimodaira-Hasegawa (SH), and Kishino-Hasegawa (KH) tests (Fig. S6). For each dataset, the alternative topologies were significantly rejected, with a *p* value less than 0.05 for all tests. The four split networks among the nine *Abies* species were consistent with the ML analyses based on the nuclear, plastid, and mitochondrial datasets (Fig. S7). In addition, the pairwise homoplasy index (PHI) test indicated that there were recombination signals in the plastid ($p = 0.002131$) and mitochondrial ($p = 0.00006591$) datasets.

To further investigate the evolutionary relationships with the genus *Abies*, 19 additional *Abies* (*A. alba*, *A. balsamea*, *A. beshanzuensis*, *A. beshanzuensis* var. *ziyuanensis*, *A. chensiensis*, *A. concolor*, *A. delavayi*, *A. delavayi* subsp. *fansipanensis*, *A. ernestii*, *A. ernestii* var. *salouenensis*, *A. fabri*, *A. fanjingshanensis*, *A. fargesii*, *A. ferreaana*, *A. forrestii*, *A. georgei* var. *smithii*, *A. nukiangensis*, *A. religiosa*, and *A. yuanbaoshanensis*) plastomes were selected (Table S9). ML analysis of the whole plastome alignment of 25 *Abies* (132,729 aligned nucleotide positions)

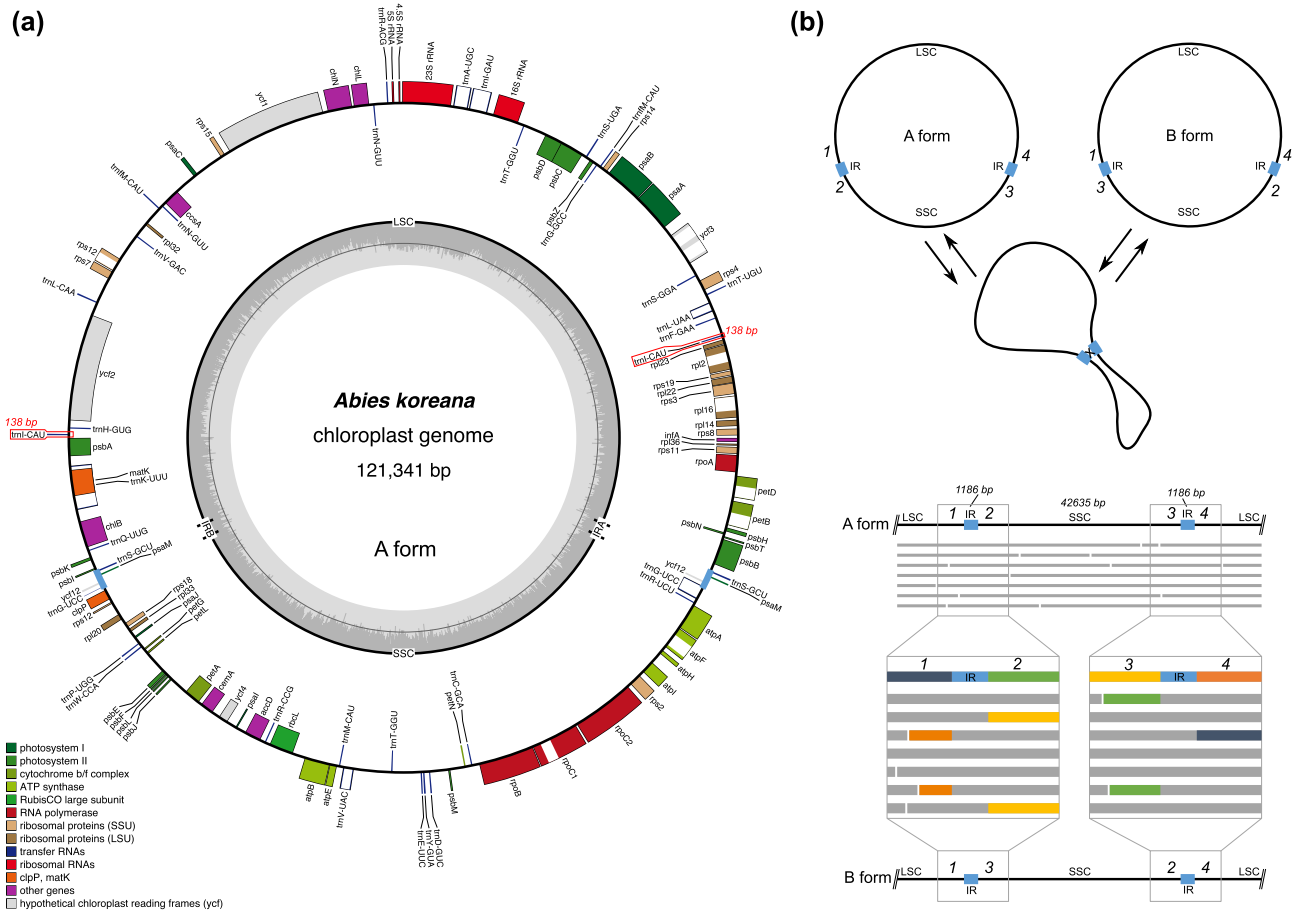


Figure 3. Plastid genome of *Abies koreana*. **(a)** Circle map of the A form. Genes on the inside and outside the map are transcribed in the clockwise and counterclockwise directions, respectively. **(b)** Formation of two isoforms. Intermolecular recombination across inverted repeats (blue). ONT reads were aligned to the plastid genome of *A. koreana*. The gray parts of the ONT reads indicate that those regions are identical to the parts of the reference genome. The dark blue, green, orange, and yellow parts of the ONT reads could not be aligned with the reference regions, which are corresponding regions. The four single-copy genomic regions surrounding the IR are shown with numbers and colors.

yielded largely congruent topologies according to the ML tree based on the whole plastid dataset of nine *Abies* (Fig. 6). The topology also showed the monophyly of *A. koreana*, *A. nephrolepis*, *A. sachalinensis*, and *A. veitchii* with bootstrap values of 100%. The topology showed the monophyly of *A. pindrow* and *A. spectabilis* with bootstrap values of 99%. The topology showed the monophyly of *A. firma* and *A. sibirica* with *A. beshanzuensis* var. *ziyuanensis* with bootstrap values of 100%. However, *A. kawakamii* nested with additional taxa and was sister to the *A. ernestii* group.

Discussion

Organelle genomes are important sources of phylogenetic information for inferring relationships within the Pinaceae family because of their different modes of inheritance, in which the mitochondrial and plastid genomes are maternally and paternally transferred, respectively. However, mitogenomes have a more complex structure (circular chromosome, linear, or complicated multibranching linear molecules) than plastomes³², making assembly difficult. As a result, 43 plastomes from 25 *Abies* taxa were sequenced. However, only one mitogenome sequence is available in GenBank (although three mitogenomes have been published). Here, we generated the complete mitogenome of *A. koreana* as a circular molecule supported by a high depth of coverage (ONT reads: mean coverage is 224x) (Fig. 1). The *A. koreana* had the smallest genome (1.17 Mb) compared to the three *Abies* mitogenomes available for *A. alba* (1.43 Mb³⁰), *A. firma* (1.33 Mb²⁸), and *A. sibirica* (1.49 Mb³¹), which were assembled into the several scaffolds (11 to 237) (Table S10). However, multiple scaffolds in the genome assembly (especially *A. firma*: 172 and *A. sibirica*: 237) can contain overlapping or redundant regions, leading to an overestimation of genome size. Incomplete or repetitive regions can also result in the creation of separate contigs or scaffolds that may actually represent the same genomic region. To obtain a more accurate estimate of *Abies* genome size, complete mitogenomes are needed. In gymnosperms, the mitogenome size varies greatly, from 346 kb in *G. biloba*²¹ to 11.69 Mb in *L. sibirica*²³. The size of the *A. koreana* mitogenome is similar to that of the *Pinus taeda* mitogenome (1.19 Mb)²⁷, which is the median among the sequenced gymnosperm mitogenomes. A comparison of the *A. koreana* mitogenome with the five representative gymnosperm mitogenomes (*Cycas*,

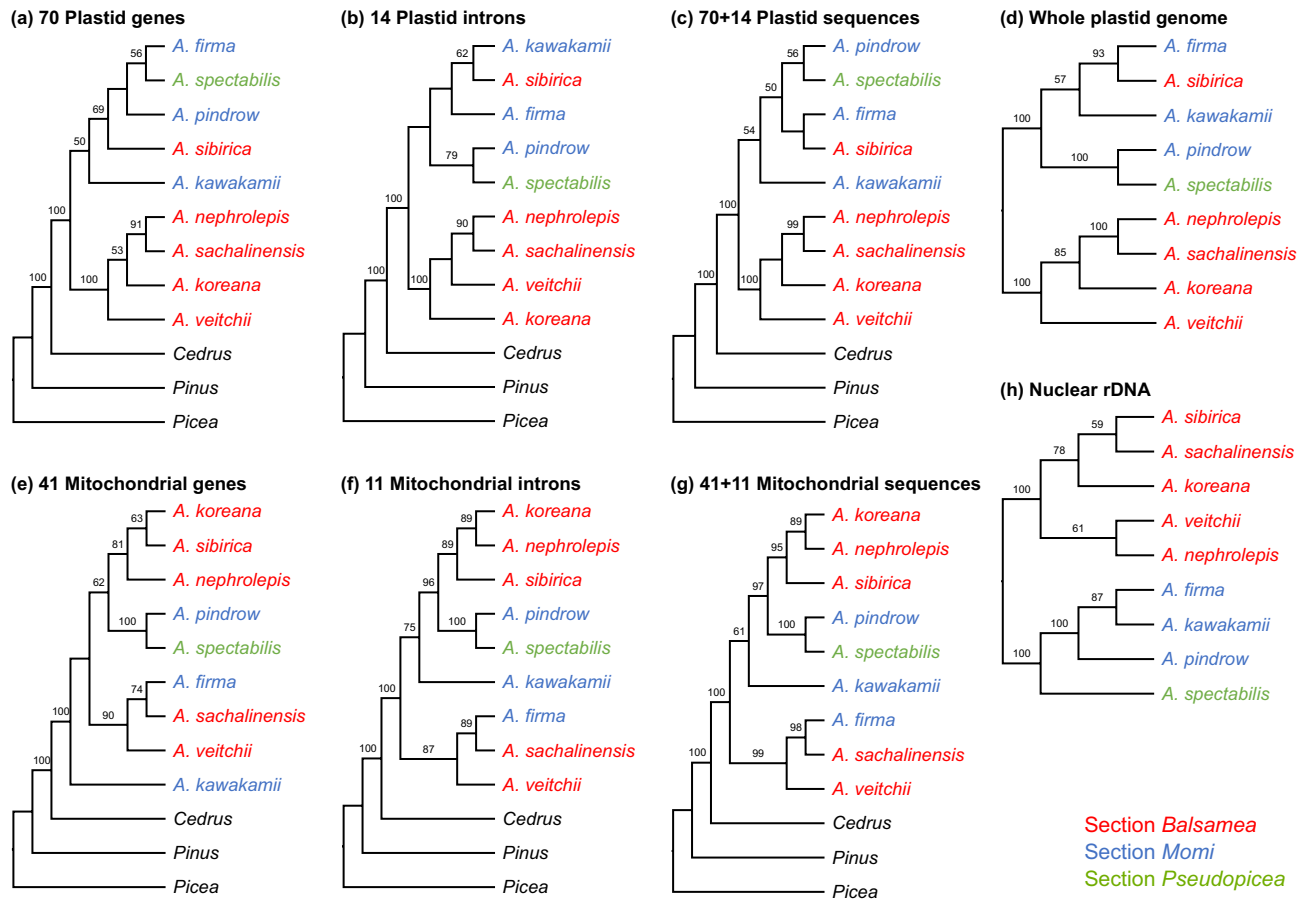


Figure 4. Phylogenetic relationships among the analyzed *Abies* species. Cladograms based on plastid (a–d), mitochondrial (e–g), and nuclear (h) sequences. Maximum likelihood bootstrap support values are shown above the branches on each cladogram.

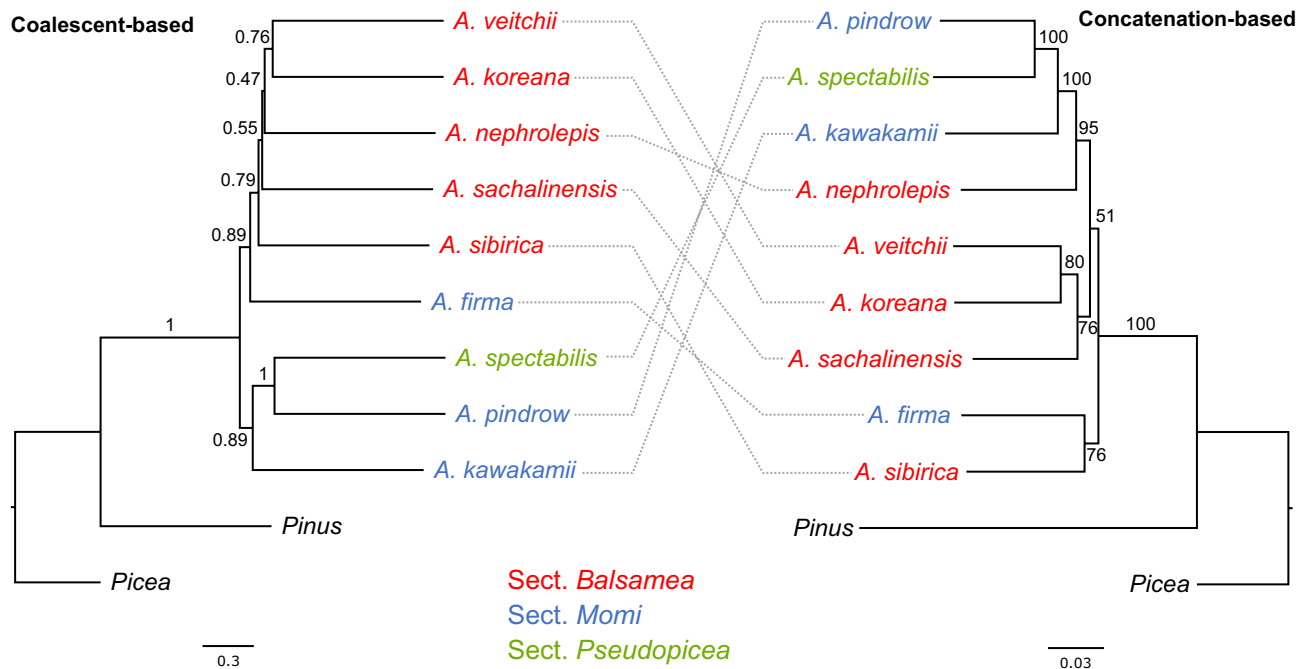


Figure 5. Phylogenomic trees of the nuclear single-copy genes inferred with ASTRAL-III and IQ-TREE2. Local posterior probability and maximum likelihood bootstrap support values are shown above the branches on each phylogram.

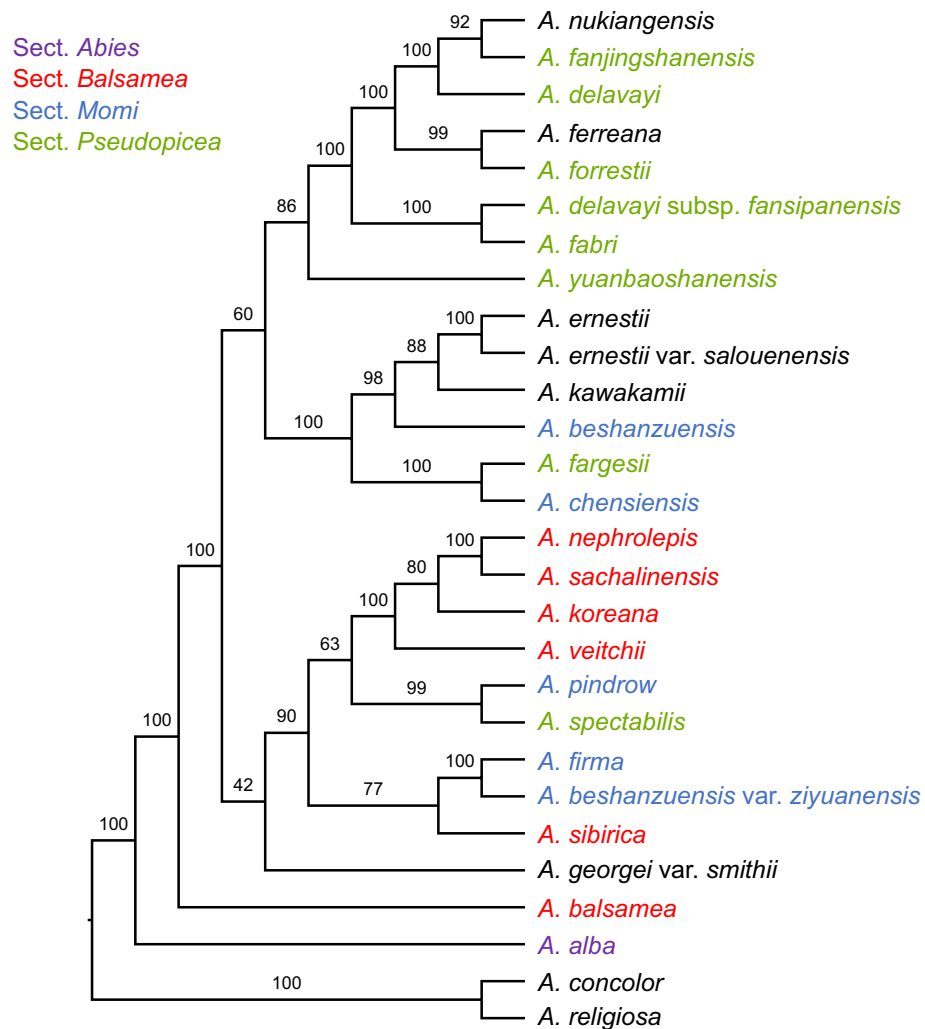


Figure 6. Phylogenetic relationships among the 25 *Abies* species based on whole plastid genomes. Maximum likelihood bootstrap support values > 50% are shown above the branches on each cladogram.

Ginkgo, *Pinus*, *Taxus*, and *Welwitschia*)²² revealed that MIPTs, MINCs, or repeats often influence variations in mitogenome size. In the *A. koreana* mitogenome, 154 kb repeats (13.1%) and 48 kb TEs (4.15%) were identified, whereas only 0.5 kb MIPTs (0.04%) were found. The amount of unidentified DNA (77.3%) contributed to the *A. koreana* mitogenome size; a nuclear genome is needed to fully understand the reasons for the mitogenome expansion.

Recombination of large repeats promotes genomic rearrangements, leading to the inference of numerous isoform maps³³. Nine large repeats (> 1 kb) may be associated with homologous recombination in the *A. koreana* mitogenome, generating multiple subgenomic circles. Recombination can disrupt conserved gene clusters in the mitogenome. We found only three gene clusters in the *A. koreana* mitogenome (Table S3). The frequency of *trans*-splicing also suggests that chromosomal rearrangements are the predominant source of transitions from *cis*- to *trans*-splicing in plant mitochondria³¹. In the *A. koreana* mitogenome, 13 of 26 introns were *trans*-splicing, which is a high number compared to that of other gymnosperm mitogenomes. This phenomenon is also observed in Pinaceae mitogenomes, suggesting that a high rate of genomic rearrangement occurred in the common ancestor of Pinaceae. Mitochondrial gene transfer to the nucleus is also an ongoing process in gymnosperms²⁸, and most of the transferred genes are ribosomal protein genes (*rps1*, *rps2*, *rps7*, *rps10*, *rps11*, *rps14*, and *rpl2*) and one succinate dehydrogenase protein gene (*sdh3*). However, our results confirmed that the *A. koreana* mitogenome also retained all 41 protein-encoding genes. Although *A. koreana* contains a full set of protein-coding genes, we performed a BLAST search between the protein and *A. koreana* transcriptomes for intracellular gene transfer (IGT) to the nucleus as an intermediate stage. Mitochondrial genes must be transferred to the nucleus before being lost from the mitogenome, but no evidence for IGT was detected in the *A. koreana* transcriptome. Taken together, the genomic rearrangement was not correlated with the gene content of mitochondrial genes in the *A. koreana* mitogenome.

Compared with RNA editing in the five mitogenomes²², *A. koreana* mitochondrial genes retained high levels of C-to-U RNA editing sites (1268 sites). We also identified mitochondrial RNA editing sites in *A. koreana* by

comparing RNA-seq raw reads and the transcriptome data, identifying 1356 C-to-U RNA editing sites in 41 mitochondrial genes. This difference occurred because PREP-MT predicted only nonsilent RNA editing sites in protein-encoding genes. In addition, these empirical data also detected start codons (*cox1*, *nad6*, *rpl2*, and *rps4*) and stop codons (*atp4*, *cox1*, *rpl2*, and *sdh3*), which were not predicted by PREP-Mt (Table S7). Cytidine-to-uridine editing can generate start or stop codons³⁴. In the *A. koreana* mitogenome, 13 of 41 genes were affected by C-to-U editing of start or stop codons. RNA-editing sites were associated with all stop codons: CAA (Q) to TAA (X) in the *atp4*, *atp6*, *ccmFc*, and *rpl16* genes; CAG (Q) to TAG (X) in the *cox1* and *rpl2* genes; and CGA (R) to TGA (X) in the *atp9*, *nad4*, *nad4L*, and *sdh3* genes. The *A. koreana* mitochondrial *rpl16* and *rps19* genes showed the GUG in the transcripts. Translation initiation at GTG and ATA codons has been proposed for a number of plant mitochondrial genes, including *apt8*, *cob*, *mttB*, *rpl16*, and *rps12*^{35–38}. These results suggest that the mitochondrial *rps19* gene of *A. koreana* can also be initiated at alternative start codons. In particular, a premature stop codon by RNA-editing in the *cox1* gene occurred at the N-terminus, followed by a new start codon. Potential functional replacement of the mitochondrial *cox1* gene by transfer to the nucleus in *A. koreana* was not detected, indicating that this gene probably encodes a functional protein in mitochondria.

The sequenced *A. koreana* plastome exhibited moderate variation in genome size and gene/intron content among the published Pinaceae plastomes. High reduction of the typical inverted repeats (IRs), which contain only *trnI-CAU* and a portion of 3' *psbA* (236–496 bp), is a well-known phenomenon in Pinaceae plastomes³⁹. The plastomes of nine *Abies* species have an extremely reduced pair of typical IRs, including only *trnI-CAU* (139 bp). In Pinaceae plastomes, multiple isoforms have been identified due to recombination at small repeats that cause genomic rearrangements^{13,40}. Within the genus *Abies*, two distinct forms (A and B) were observed depending on the species, and only the A form of *A. koreana* was detected by PCR¹³. However, for the *A. koreana* plastome reported here, four independent assemblies using short and long reads verified that there were alternative isoforms. Mapping of the long reads to the A form confirmed that *A. koreana* has two isoforms, which are achieved by homologous recombination with a pair of short repeats (1186 bp) (Fig. 3b).

The nuclear, plastid and mitochondrial datasets analyzed in this study (Figs. 4 and 5) represent the most extensive sets of genes from the three plant genomes for *Abies* phylogenetics. Our analyses were largely consistent with previous studies based on the nuclear ITS region³, five plastid genes, and a single nuclear locus⁴; and a combined analysis of six mitochondrial, plastid, and nuclear ITS regions²; and large portion of the mitochondrial genes⁹. Collectively, these studies demonstrate that *Abies* is monophyletic and sister to the genus *Keteleeria*. The extended phylogeny presented here (Fig. 6) is strongly inconsistent with the traditional classification⁴¹. Our analyses also suggested high levels of genome-tree conflict between nuclear and cytoplasmic genomes or between organellar genomes, although not all sections of the genus *Abies* were sampled in this study. The conflict between plastid and mitochondrial genomes is possibly due to the different inheritance of organelles in the genus *Abies*. The main causes of cytonuclear discordance are hybridization or introgression and incomplete lineage sorting. Previous studies and the finding of the present study suggest that recurrent hybridization and introgression events may have contributed to the evolutionary history of eastern Asian *Abies*^{8,42}. Thus, these phenomena in the genus *Abies* occur via organellar genome capture⁴³.

Combining mitochondrial protein-coding genes with introns, rather than only using mitochondrial protein-coding genes alone, is more advantageous for generating an abundance of variable characters (Fig. 4). Additionally, the use of nuclear 5'-ETS region sequences for phylogenetic analysis has improved resolution. However, the presence of ITS polymorphisms has been observed in Pinaceae, which may be a result of a slow rate of concerted evolution⁴⁴.

Given the incongruence among genomes for *Abies* relationships, we urge caution in combining datasets derived from different genomes to study the phylogenetics of the genus *Abies*. A full understanding of the complicated evolutionary history of the genus *Abies* requires extensive sampling of all three plant genomes (especially nuclear transcriptomes) from all sections.

Conclusions

The complete mitogenome of *A. koreana* provides important additional information for improving the understanding of genome evolution among gymnosperms in terms of plastid-derived DNA sequences, repeats, and RNA editing sites. Our results also provide new insight into the structural variation in the *A. koreana* plastome. Based on the organelle genome-scale data with nrDNA, phylogenomic relationships among the nine analyzed *Abies* species were strongly supported, showing conflicting topologies. Overall, our research has greatly enriched the genome resources of *Abies*, which will help to further analyze the phylogeny and suggest new approaches for elucidating the evolutionary relationships within *Abies* in the future.

Methods

DNA isolation and sequencing

Fresh leaves of *A. koreana* were collected from a single tree in the garden of the National Institute of Biological Resources (NIBR, Incheon, South Korea) and a voucher specimen was deposited in the NIBR Herbarium (NIBRVP0000729725 identified by Myounghai Kwak). The experimental study of the plant, including collection of the material, complied with institutional, national, and international guidelines. Total genomic DNA was extracted from the dried bulbous by grinding them with liquid nitrogen using the cetyltrimethylammonium bromide (CTAB) method⁴⁵ and then sequenced to generate 100-bp × 2 paired-end (PE) reads from a 550-bp library using the Illumina NovaSeq 6000 platform (Illumina, San Diego, CA). A single flow cell of long reads was sequenced using the Oxford Nanopore Technologies PromethION platform (ONT, Oxford, United Kingdom). Additional gDNA from seven *Abies* species (*A. kawakamii*, *A. nephrolepis*, *A. pindrow*, *A. sachalinensis*, *A. sibirica*,

A. spectabilis, and *A. veitchii*) was extracted and sequenced to generate 150-bp × 2 PE reads using the Illumina NovaSeq 6000 platform.

Genome assembly, annotations, and analyses

The *A. koreana* plastid and mitochondrial genomes were generated from contigs produced by Velvet v1.2.10⁴⁶, SPAdes v3.13.1⁴⁷, and MaSuRCA v3.2.6⁴⁸. For Velvet assemblies, each run was performed with only PE reads using multiple *k*-mers (69 to 93) and expected coverage values (50, 100, 200, 500, and 1000). For SPAdes assemblies, the coverage cutoff was set to 20, 50, and 100 with mismatch correction mode (-careful) using both PE and ONT reads. For MaSuRCA assembly, a configuration file containing default assembly parameters was created with both PE and ONT reads. Illumina data for the seven *Abies* species were assembled with Velvet as described above for *A. koreana*. For each species, nuclear ribosomal DNA (nrDNA), plastid, and mitochondrial contigs were identified using a BLAST-like algorithm in Geneious Prime 2022.0.1 (<http://www.geneious.com>) with the organelle genes of *Pinus taeda* plastid (NC_021440) and mitochondrial (NC_039746) genomes as queries. The consensus nrDNA, plastid, and mitochondrial genome sequences were completed by manually aligning the overlapping contigs. Genome sequencing data were obtained from the NCBI Sequence Read Archive (SRA) for one additional *Abies* species (*A. firma*; SRR12710828) and three additional Pinaceae species (*Cedrus deodara*; SRR12710827, *Picea smithiana*; SRR12710829, and *Pinus armandii*; SRR12710830), and the nrDNA and plastid genome sequences were recovered as described above.

Circular plastid and mitochondrial genome maps were drawn using OrganellarGenomeDRAW (OGDRAW) v1.3.1 (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>)⁴⁹. The newly generated plastid, mitochondrial, and nrDNA sequences were deposited in GenBank (Table S8) and the Dryad Digital Repository.

Using ROUSFinder.py⁵⁰, repetitive sequences in the *A. koreana* mitochondrial and plastid genomes were discovered. MIPT sequences were identified by performing BLASTN searches of the *A. koreana* plastid genome against the mitochondrial genome in Geneious Prime with an e-value cutoff of 1e-10, at least 80% sequence identity and a minimum length of 50 bp. The "Find ORFs" with an ATG start codon in Geneious Prime were used to find open reading frames (ORFs) in the mitochondrial genome of *A. koreana* that were longer than 150 bp. To search for chimeric ORFs, all ORFs were compared with annotated *A. koreana* mitochondrial genes using BLASTN with an e-value cutoff of 1e-3, minimum length of 30 bp (as described in Mower et al.⁵¹) and at least 90% sequence identity. Using the TMHMM server v.2.0⁵², transmembrane helices in the selected ORFs were predicted.

RNA editing sites for mitochondrial genes were predicted using PREP-Mt⁵³ with a cutoff value of 0.2. To check the empirical RNA editing sites on the protein-coding genes, paired-end Illumina sequence reads were also downloaded from the NCBI SRA repository (SRR5747840). We performed error correction for the raw reads using Rcorrector v1.0.4⁵⁴ and then mapped the corrected reads to the genomic gene sequences using Bowtie v2.2.9⁵⁵. In addition, the transcriptome was assembled by Trinity⁵⁶ with the "trimmomatic" option using the corrected reads. The available mitochondrial transcripts in the transcriptome were identified using BLAST + v2.12.0⁵⁷ and then aligned with the genomic gene sequences.

Phylogenetic analyses

Each organelle gene was aligned based on the back-translation approach with MAFFT v7.450⁵⁸ in Geneious Prime 2022.0.1. Nuclear rDNA sequences (5'-ETS, 18S, ITS1, 5.8S, ITS2, and 26S) were aligned using MUSCLE v3.8.425⁵⁹ in Geneious Prime. To recover nuclear single-copy gene sequences from the PE reads, an alignment-based method was employed using Read2Tree v0.1.5⁶⁰. A total of 1000 orthologous markers for *Picea glauca* and *Pinus taeda* were obtained from the OMA browser (https://omabrowser.org/oma/export_markers)⁶¹. Read2Tree was utilized to generate nucleotide sequence alignments using these marker genes. Phylogenetic analyses of each dataset were performed using IQ-TREE2 v2.1.4-beta⁶² with a best-fit model (-m TEST) and 1000 ultrafast bootstrap replicates. ASTRAL v5.7.8⁶³ was used for the datasets of 452 nuclear genes. Neighbor-Net was inferred using SplitsTree v4.15.1⁶⁴ with uncorrected *p*-distances and 1000 bootstrap replicates. Three tree topologies were statistically tested for incongruence using approximately unbiased (AU), Shimodaira-Hasegawa (SH), and Kishino-Hasegawa (KH) methods. IQ-TREE2 estimated site likelihoods for each topology on three concatenated datasets and performed AU, SH, and KH tests. Species-level taxonomy follows that of Farjon and Rushforth⁴¹, except for the recognition of *A. kawakamii*, which is placed in Sect. *Momi* as suggested by phylogenetic studies^{3,65}.

Ethics approval and consent to participate

The experimental studies on the plants, including collection of the material, complied with institutional, national, and international guidelines.

Data availability

The datasets supporting the results of this article are included in additional files. Complete plastid and mitochondrial genome sequences are available in GenBank (<https://www.ncbi.nlm.nih.gov/nuccore/ON897689>, ON897690). DNA sequencing files are available in the NCBI Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra>) under BioProject PRJNA1088239. The phylogenetic datasets generated during the current study are available in the Dryad Digital Repository, (<https://doi.org/https://doi.org/10.5061/dryad.83bk3j9wh>).

Received: 10 April 2023; Accepted: 27 March 2024

Published online: 01 April 2024

References

- Liu, T. *A Monograph of the Genus Abies* (College of Agriculture, National Taiwan University, 1971).
- Xiang, Q.-P. *et al.* Phylogenetic relationships, possible ancient hybridization, and biogeographic history of *Abies* (Pinaceae) based on data from nuclear, plastid, and mitochondrial genomes. *Mol. Phylogenet. Evol.* **82**, 1–14. <https://doi.org/10.1016/j.ympev.2014.10.008> (2015).
- Xiang, Q.-P., Xiang, Q.-Y., Guo, Y.-Y. & Zhang, X.-C. Phylogeny of *Abies* (Pinaceae) inferred from nrITS sequence data. *Taxon* **58**, 141–152. <https://doi.org/10.1002/tax.581015> (2009).
- Xiang, Q.-P., Wei, R., Zhu, Y.-M., Harris, A. & Zhang, X.-C. New infrageneric classification of *Abies* in light of molecular phylogeny and high diversity in western North America. *J. Syst. Evol.* **56**, 562–572. <https://doi.org/10.1111/jse.12458> (2018).
- Aguirre-Planter, E. *et al.* Phylogeny, diversification rates and species boundaries of Mesoamerican firs (*Abies*, Pinaceae) in a genus-wide context. *Mol. Phylogenet. Evol.* **62**, 263–274. <https://doi.org/10.1016/j.ympev.2011.09.021> (2012).
- Semerikova, S. A., Khrunyk, Y. Y., Lascoux, M. & Semerikov, V. L. From America to Eurasia: A multigenomes history of the genus *Abies*. *Mol. Phylogenet. Evol.* **125**, 14–28. <https://doi.org/10.1016/j.ympev.2018.03.009> (2018).
- Liepert, S., Mayland-Quellhorst, E., Lahme, M. & Ziegenhagen, B. Contrasting geographical patterns of ancient and modern genetic lineages in Mediterranean *Abies* species. *Plant Syst. Evol.* **284**, 141–151. <https://doi.org/10.1007/s00606-009-0247-8> (2010).
- Peng, Y. *et al.* Phylogeographic analysis of the fir species in southern China suggests complex origin and genetic admixture. *Ann. For. Sci.* **69**, 409–416. <https://doi.org/10.1007/s13595-011-0170-3> (2012).
- Semerikov, V. L., Semerikova, S. A., Khrunyk, Y. Y. & Putintseva, Y. A. Sequence capture of mitochondrial genome with PCR-generated baits provides new insights into the biogeography of the genus *Abies* mill. *Plants* **11**, 762 (2022).
- Christenhusz, M. J. & Byng, J. W. The number of known plants species in the world and its annual increase. *Phytotaxa* **261**, 201–217 (2016).
- Mogensen, H. L. INVITED SPECIAL PAPER: The hows and whys of cytoplasmic inheritance in seed plants. *Am. J. Bot.* **83**, 383–404. <https://doi.org/10.1002/j.1537-2197.1996.tb12718.x> (1996).
- Zhong, Z.-R., Li, N., Qian, D., Jin, J.-H. & Chen, T. Maternal inheritance of plastids and mitochondria in *Cycas* L. (Cycadaceae). *Mol. Genet. Genom.* **286**, 411–416. <https://doi.org/10.1007/s00438-011-0653-9> (2011).
- Wu, C.-S., Lin, C.-P., Hsu, C.-Y., Wang, R.-J. & Chaw, S.-M. Comparative chloroplast genomes of pinaceae: Insights into the mechanism of diversified genomic organizations. *Genome Biol. Evol.* **3**, 309–319. <https://doi.org/10.1093/gbe/evr026> (2011).
- Qu, X.-J., Wu, C.-S., Chaw, S.-M. & Yi, T.-S. Insights into the existence of isomeric plastomes in Cupressoidae (Cupressaceae). *Genome Biol. Evol.* **9**, 1110–1119. <https://doi.org/10.1093/gbe/evx071> (2017).
- Wu, C.-S. & Chaw, S.-M. Evolutionary stasis in cycad plastomes and the first case of plastome GC-biased gene conversion. *Genome Biol. Evol.* **7**, 2000–2009. <https://doi.org/10.1093/gbe/evv125> (2015).
- Wu, C.-S., Lai, Y.-T., Lin, C.-P., Wang, Y.-N. & Chaw, S.-M. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: Selection toward a lower-cost strategy. *Mol. Phylogenet. Evol.* **52**, 115–124. <https://doi.org/10.1016/j.ympev.2008.12.026> (2009).
- McCoy, S. R., Kuehl, J. V., Boore, J. L. & Raubeson, L. A. The complete plastid genome sequence of *Welwitschia mirabilis*: An unusually compact plastome with accelerated divergence rates. *BMC Evol. Biol.* **8**, 130. <https://doi.org/10.1186/1471-2148-8-130> (2008).
- Lin, C.-P., Huang, J.-P., Wu, C.-S., Hsu, C.-Y. & Chaw, S.-M. Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. *Genome Biol. Evol.* **2**, 504–517. <https://doi.org/10.1093/gbe/evq036> (2010).
- Chaw, S.-M. *et al.* The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, *bpu* sequences, and abundant RNA editing sites. *Mol. Biol. Evol.* **25**, 603–615. <https://doi.org/10.1093/molbev/msn009> (2008).
- Habib, S., Dong, S., Liu, Y., Liao, W. & Zhang, S. The complete mitochondrial genome of *Cycas debaoensis* revealed unexpected static evolution in gymnosperm species. *PLoS ONE* **16**, e0255091. <https://doi.org/10.1371/journal.pone.0255091> (2021).
- Guo, W. *et al.* *Ginkgo* and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Mol. Biol. Evol.* **33**, 1448–1460. <https://doi.org/10.1093/molbev/msw024> (2016).
- Kan, S.-L., Shen, T.-T., Gong, P., Ran, J.-H. & Wang, X.-Q. The complete mitochondrial genome of *Taxus cuspidata* (Taxaceae): Eight protein-coding genes have transferred to the nuclear genome. *BMC Evol. Biol.* **20**, 10. <https://doi.org/10.1186/s12862-020-1582-1> (2020).
- Putintseva, Y. A. *et al.* Siberian larch (*Larix sibirica* Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. *BMC Genomics* **21**, 654. <https://doi.org/10.1186/s12864-020-07061-4> (2020).
- Sullivan, A. R. *et al.* The mitogenome of Norway spruce and a reappraisal of mitochondrial recombination in plants. *Genome Biol. Evol.* **12**, 3586–3598. <https://doi.org/10.1093/gbe/evz263> (2019).
- Jackman, S. D. *et al.* Complete mitochondrial genome of a gymnosperm, Sitka Spruce (*Picea sitchensis*), indicates a complex physical structure. *Genome Biol. Evol.* **12**, 1174–1179. <https://doi.org/10.1093/gbe/evaa108> (2020).
- Jackman, S. D. *et al.* Organellar genomes of White Spruce (*Picea glauca*): Assembly and annotation. *Genome Biol. Evol.* **8**, 29–41. <https://doi.org/10.1093/gbe/evv244> (2015).
- Neale, D. B. *et al.* Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol.* **15**, R59. <https://doi.org/10.1186/gb-2014-15-3-r59> (2014).
- Kan, S.-L., Shen, T.-T., Ran, J.-H. & Wang, X.-Q. Both Conifer II and Gnetales are characterized by a high frequency of ancient mitochondrial gene transfer to the nuclear genome. *BMC Biol.* **19**, 146. <https://doi.org/10.1186/s12915-021-01096-z> (2021).
- Wu, C.-S. & Chaw, S.-M. Evolution of mitochondrial RNA editing in extant gymnosperms. *Plant J.* **111**, 1676–1687. <https://doi.org/10.1111/tpj.15916> (2022).
- Kersten, B. *et al.* The mitochondrial genome sequence of *Abies alba* Mill. reveals a high structural and combinatorial variation. *BMC Genomics* **23**, 776. <https://doi.org/10.1186/s12864-022-08993-9> (2022).
- Guo, W., Zhu, A., Fan, W., Adams, R. P. & Mower, J. P. Extensive shifts from *Cis*- to *Trans*-splicing of gymnosperm mitochondrial introns. *Mol. Biol. Evol.* **37**, 1615–1620. <https://doi.org/10.1093/molbev/msaa029> (2020).
- Sloan, D. B. One ring to rule them all? Genome sequencing provides new insights into the ‘master circle’ model of plant mitochondrial DNA structure. *New Phytol.* **200**, 978–985. <https://doi.org/10.1111/nph.12395> (2013).
- Gualberto, J. M. & Newton, K. J. Plant mitochondrial genomes: Dynamics and mechanisms of mutation. *Annu. Rev. Plant Biol.* **68**, 225–252 (2017).
- Gott, J. M. & Emeson, R. B. Functions and mechanisms of RNA editing. *Annu. Rev. Genet.* **34**, 499–531. <https://doi.org/10.1146/annurev.genet.34.1.499> (2000).
- Bock, H., Brennicke, A. & Schuster, W. *Rps3* and *rpl16* genes do not overlap in *Oenothera* mitochondria: GTG as a potential translation initiation codon in plant mitochondria?. *Plant Mol. Biol.* **24**, 811–818 (1994).
- Sünkel, S., Brennicke, A. & Knoop, V. RNA editing of a conserved reading frame in plant mitochondria increases its similarity to two overlapping reading frames in *Escherichia coli*. *Mol. Gen. Genet. MGG* **242**, 65–72 (1994).
- Siculella, L., Pacoda, D., Treglia, S., Gallerani, R. & Ceci, L. GTG as translation initiation codon in the apocytochrome b gene of sunflower mitochondria. *DNA Sequence* **6**, 365–369 (1996).

38. Zhu, A., Guo, W., Jain, K. & Mower, J. P. Unprecedented heterogeneity in the synonymous substitution rate within a plant genome. *Mol. Biol. Evol.* **31**, 1228–1236. <https://doi.org/10.1093/molbev/msu079> (2014).
39. Chaw, S.-M., Wu, C.-S. & Sudianto, E. In *Adv. Bot. Res.* Vol. 85 (eds Chaw, S.-M. & Jansen, R. K.) 195–222 (Academic Press, 2018).
40. Tsumura, Y., Suyama, Y. & Yoshimura, K. Chloroplast DNA inversion polymorphism in populations of *Abies* and *Tsuga*. *Mol. Biol. Evol.* **17**, 1302–1312. <https://doi.org/10.1093/oxfordjournals.molbev.a026414> (2000).
41. Farjon, A. & Rushforth, K. A classification of *Abies miller* (Pinaceae). *Misc. Publ. Univ. Utrecht Herb.* **3**, 59–79 (1988).
42. Tsumura, Y. & Suyama, Y. Differentiation of mitochondrial DNA polymorphisms in populations of five Japanese *Abies* species. *Evolution* **52**, 1031–1042. <https://doi.org/10.1111/j.1558-5646.1998.tb01831.x> (1998).
43. Folk, R. A., Mandel, J. R. & Freudenstein, J. V. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Syst. Biol.* **66**, 320–337. <https://doi.org/10.1093/sysbio/syw083> (2016).
44. Gernandt, D. S. & Liston, A. Internal transcribed spacer region evolution in *Larix* and *Pseudotsuga* (Pinaceae). *Am. J. Bot.* **86**, 711–723 (1999).
45. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. (1987).
46. Zerbino, D. R. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829. <https://doi.org/10.1101/gr.074492.107> (2008).
47. Antipov, D., Korobeynikov, A., McLean, J. S. & Pevzner, P. A. hybridSPAdes: An algorithm for hybrid assembly of short and long reads. *Bioinformatics* **32**, 1009–1015. <https://doi.org/10.1093/bioinformatics/btv688> (2016).
48. Zimin, A. V. et al. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res.* **27**, 787–792. <https://doi.org/10.1101/gr.213405.116> (2017).
49. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **47**, W59–W64. <https://doi.org/10.1093/nar/gkz238> (2019).
50. Wynn, E. L. & Christensen, A. C. Repeats of unusual size in plant mitochondrial genomes: Identification, incidence and evolution. *G3 Genes Genomes Genet.* **9**, 549–559 (2019).
51. Mower, J. P., Case, A. L., Floro, E. R. & Willis, J. H. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. *Genome Biol. Evol.* **4**, 670–686. <https://doi.org/10.1093/gbe/evs042> (2012).
52. Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
53. Mower, J. P. PREP-Mt: Predictive RNA editor for plant mitochondrial genes. *BMC Bioinform.* **6**, 96. <https://doi.org/10.1186/1471-2105-6-96> (2005).
54. Song, L. & Florea, L. Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* **4**, 48. <https://doi.org/10.1186/s13742-015-0089-y> (2015).
55. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359. <https://doi.org/10.1038/nmeth.1923> (2012).
56. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652. <https://doi.org/10.1038/nbt.1883> (2011).
57. Camacho, C. et al. BLAST+: Architecture and applications. *BMC Bioinform.* **10**, 421. <https://doi.org/10.1186/1471-2105-10-421> (2009).
58. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. <https://doi.org/10.1093/molbev/mst010> (2013).
59. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797. <https://doi.org/10.1093/nar/gkh340> (2004).
60. Dylus, D., Altenhoff, A., Majidian, S., Sedlazeck, F. J. & Dessimoz, C. Inference of phylogenetic trees directly from raw sequencing reads using Read2Tree. *Nat. Biotechnol.* **42**, 139–147. <https://doi.org/10.1038/s41587-023-01753-4> (2024).
61. Altenhoff, A. M. et al. OMA orthology in 2024: Improved prokaryote coverage, ancestral and extant GO enrichment, a revamped synteny viewer and more in the OMA Ecosystem. *Nucleic Acids Res.* **52**, D513–D521. <https://doi.org/10.1093/nar/gkad1020> (2023).
62. Minh, B. Q. et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534. <https://doi.org/10.1093/molbev/msaa015> (2020).
63. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform.* **19**, 153. <https://doi.org/10.1186/s12859-018-2129-y> (2018).
64. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
65. Xiang, Q.-P., Xiang, Q.-Y., Liston, A. & Zhang, X.-C. Phylogenetic relationships in *Abies* (Pinaceae): Evidence from PCR-RFLP of the nuclear ribosomal DNA internal transcribed spacer region. *Bot. J. Linn. Soc.* **145**, 425–435. <https://doi.org/10.1111/j.1095-8339.2004.00286.x> (2004).

Author contributions

SP contributed to the design of the project, assembled the genomes, performed the analyses, prepared the figures, and read/edited the manuscript. MK and SJP contributed to the project design and read/edited the manuscript. All the authors have read and approved the final draft of the manuscript.

Funding

This research was supported by the National Institute of Biological Resources Grant (NIBR201803102 to MK), South Korea.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-58253-x>.

Correspondence and requests for materials should be addressed to M.K. or S.P.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024