



OPEN

# Large-scale phylogenomics reveals ancient introgression in Asian *Hepatica* and new insights into the origin of the insular endemic *Hepatica maxima*

Seongjun Park<sup>1,2</sup> & SeonJoo Park<sup>2</sup>✉

*Hepatica maxima* is native to Ulleungdo, which is one of the oceanic islands in Korea, and it likely originated via anagenetic speciation from the Korean mainland species *H. asiatica*. However, the relationships among the Asian lineages remain unresolved. Phylogenomics based on plant genomes can provide new insights into the evolutionary history of plants. We first generated plastid, mitochondrial and transcriptome sequences of the insular endemic species *H. maxima*. Using the genomic data for *H. maxima*, we obtained a phylogenomic dataset consisting of 76 plastid, 37 mitochondrial and 413 nuclear genes from Asian *Hepatica* and two outgroups. Coalescent- and concatenation-based methods revealed cytonuclear and organellar discordance in the lineage. The presence of gynodioecy with cytoplasmic male sterility in Asian *Hepatica* suggests that the discordance is correlated with potential disruption of linkage disequilibrium between the organellar genomes. Species network analyses revealed a deep history of hybridization and introgression in Asian *Hepatica*. We discovered that ancient and recent introgression events occurred throughout the evolutionary history of the insular endemic species *H. maxima*. The introgression may serve as an important source of genetic variation to facilitate adaptation to the Ulleungdo environment.

Oceanic islands are ideal evolutionary laboratories for investigating the processes of plant speciation. Two modes of speciation, cladogenesis and anagenesis, are observed on oceanic islands<sup>1</sup>. During cladogenesis, an initial population splits into two lineages through isolation, which leads to genetic and morphological variation between the lineages<sup>1</sup>. During anagenesis, continental progenitor populations slowly transform via genetic and morphological variation through time without a lineage-splitting event<sup>1–3</sup>. Oceanic islands are usually small land masses that are well isolated from the mainland and have high levels of endemism<sup>1</sup>. Ulleungdo is an oceanic island with a total area of 73 km<sup>2</sup>, and it is located 137 km east of the Korean Peninsula<sup>4</sup>. It is home to a small flora of only 33 endemic plant species<sup>2,5</sup>. However, Ulleungdo provides great opportunities for understanding the two principal modes of speciation and shows a higher level of endemism derived via anagenesis (88%) than cladogenesis (12%)<sup>2</sup>. Multiple studies have focused on understanding the genetic variation between endemic species on Ulleungdo and their continental progenitors<sup>6–9</sup>.

The genus *Hepatica* (Ranunculaceae) is distributed in temperate zones of the Northern Hemisphere<sup>10</sup> and shows high diversification in eastern Asia characterized by four species (*H. asiatica*, *H. henryi*, *H. insularis*, and *H. maxima*) and two varieties (*H. nobilis* var. *japonica* and *H. nobilis* var. *pubescens*). *Hepatica maxima* (Nakai) Nakai is endemic to Ulleungdo and likely originated via anagenetic speciation<sup>11</sup> from the Korean mainland species *H. asiatica*<sup>12</sup>. *Hepatica maxima* exhibits “insular gigantism”, with increased leaf, bract and seed sizes compared to those of the putative progenitor<sup>13</sup>. In addition, *H. maxima* is distinguished by glabrous achenes and leaf cuticles<sup>13,14</sup>. A previous study estimated the phylogenetic relationships within the genus *Hepatica* using nuclear (ITS) and plastid (*atpB-rbcL* spacer and *trnL-F* intron/spacer) markers and amplified fragment length

<sup>1</sup>Institute of Natural Science, Yeungnam University, Gyeongsan, Gyeongbuk 38541, South Korea. <sup>2</sup>Department of Life Sciences, Yeungnam University, Gyeongsan, Gyeongbuk 38541, South Korea. ✉email: sjpark01@ynu.ac.kr

polymorphism (AFLP) data<sup>12</sup>, however the relationships within the genus *Hepatica* remain unresolved. Resolving interspecific relationships within *Hepatica* is important for reconstructing the evolutionary history of *H. maxima*.

Plant cells have three distinct genomes (nuclear, plastid and mitochondrial genomes), with each providing valuable resources for studying plant evolution. For example, organellar genomic information provides a better understanding of plant evolution<sup>15,16</sup>. In particular, variation in organellar genomes can contribute to plant adaptation<sup>17</sup>. Plastid and mitochondrial genomes (the plastome and mitogenome, respectively) are generally inherited maternally in higher plants<sup>18,19</sup>, although several exceptions are observed<sup>20,21</sup>. Compared with organellar genomes, the nuclear genome is biparentally inherited and presents better resolution among closely related species with high mutation rates<sup>22,23</sup>. Nuclear genes provide powerful opportunities to address adaptation to environmental changes<sup>24–27</sup>.

Genome-scale data improve the understanding of the origin of species and lineages that undergo rapid evolutionary radiation. Thus, comparative analyses of nuclear, plastid, and mitochondrial genomes are required for understanding the origin of the insular endemic species *H. maxima*. The development of next-generation sequencing (NGS) techniques and methods<sup>28,29</sup> enables us to assemble complete plastomes and mitogenomes as well as nuclear transcriptomes for plants. In addition, third-generation sequencing platforms such as those of Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT)<sup>30</sup> facilitate the assembly of large and complex genomes. However, while one plastome sequence is available for *Hepatica*, data on the mitochondrial and nuclear genomes in this genus is lacking.

In this study, we first generated the plastid and mitochondrial genomes and nuclear transcriptome of the insular endemic species *H. maxima*. To reconstruct a genome-scale phylogeny for the Asian *Hepatica*, we sequenced complete plastome and draft mitogenome sequences from six additional taxa, including four *Hepatica* as well as *Anemone narcissiflora* and *Pulsatilla koreana*, as well as six nuclear transcriptomes. The patterns of incongruence between organellar and nuclear genomes and within the organellar genomes were identified. Additional comparative genomic and phylogenomic analyses provide new insights into organellar genome evolution within tribe Anemoneae and the evolutionary origin of the insular endemic species *H. maxima*.

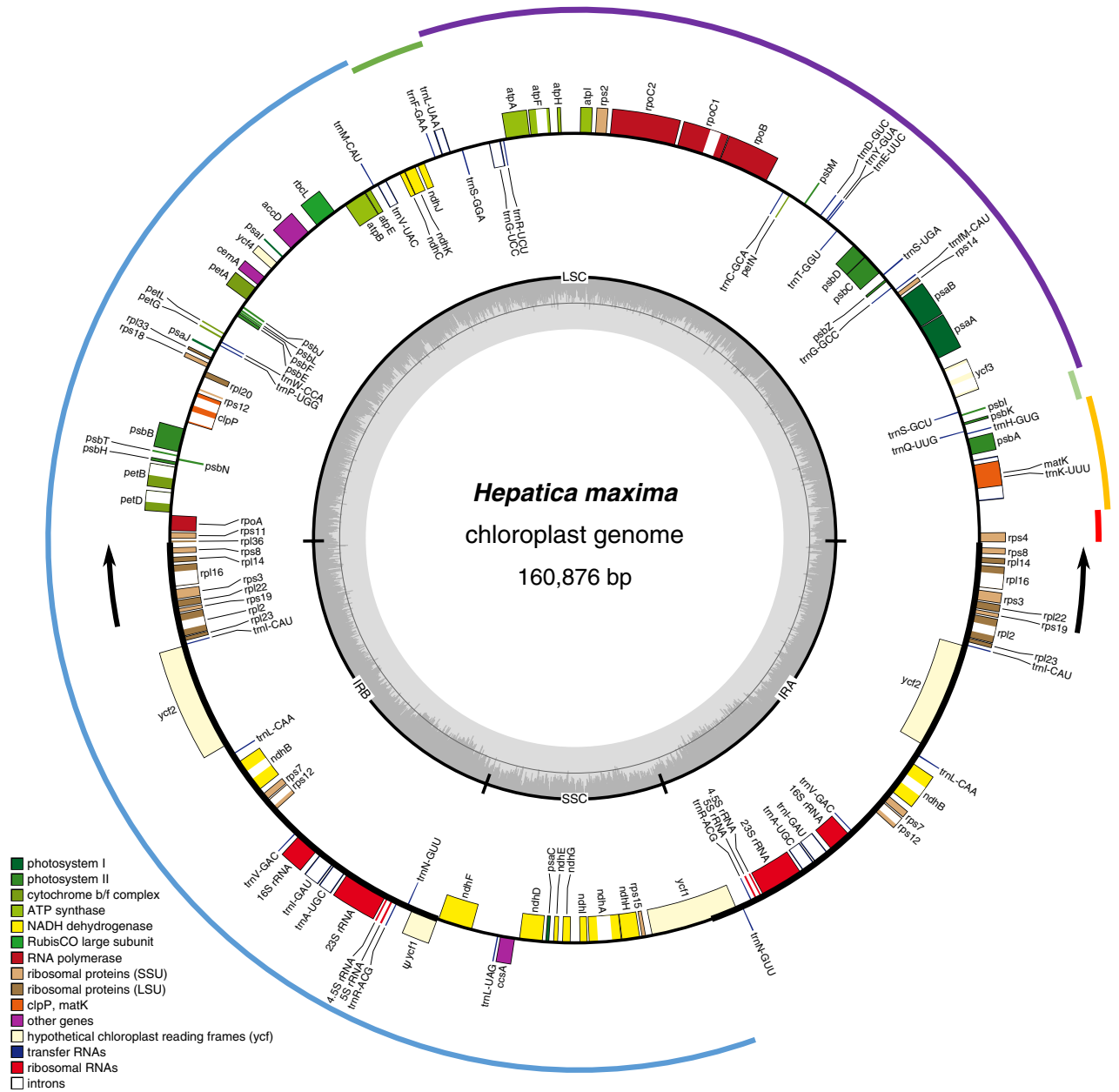
## Results

**Plastid and mitochondrial genomes.** The organellar genomes of the insular endemic species *H. maxima* were sequenced on the Illumina and Oxford platforms and completed by Velvet and Spades assemblies with optimized parameter values. The plastome of *H. maxima* was 160,876 bp in length ( $\approx 1250 \times$  depth of coverage), with a pair of inverted repeats (IRs) (IR<sub>A</sub> and IR<sub>B</sub>) of 31,097 bp separated by large and small single-copy (LSC and SSC) regions of 80,998 and 17,684 bp, respectively (Fig. 1 and Table 1). The plastome contained 110 different genes, including 76 protein-coding genes, 29 tRNA genes, and four rRNA genes (Table 1). Three protein-coding genes, namely, translation initiation factor A (*infA*), ribosomal protein subunit L32 (*rpl32*), and ribosomal protein subunit S16 (*rps16*), and one tRNA gene, *trnT-UGU*, were absent. The plastome had an expanded IR with the IR<sub>A</sub>/LSC boundary shifted from *rps19* into *rps8* (Fig. 1).

The *H. maxima* mitogenome was assembled into a 1,122,546 bp circular molecule ( $\approx 110 \times$  depth of coverage) and contained 39 protein-coding genes, 3 rRNA genes, 11 tRNA genes, 18 *cis*-spliced introns, and 5 *trans*-spliced introns (Table 1 and Fig. 2). Two ribosomal protein genes, *rps10* and *rps14*, were absent or appeared to be pseudogenes. The mitogenome contained 77,887 bp of repetitive DNA, ranging from 31 to 19,922 bp in length. The mitogenome possessed 23 fragments of plastid-derived sequences ranging from 51 bp to 7,104 bp in length (Supplementary Table S1) and covering 1.54% of the genome. PREP-Mt predicted 687 putative C-to-U RNA editing sites in the 39 mitochondrial protein-coding genes (Supplementary Table S2). Eighteen open reading frames (ORFs) at least 150 bp in length were identified that contain small fragments ( $> 30$  bp) of one or more mitochondrial genes (Supplementary Table S3). These ORFs contained small fragments of the nine mitochondrial genes *atp1*, *atp6*, *atp9*, *cox1*, *ccmFc*, *nad6*, *rps19*, *rps4*, and *rrn26*. Five of these ORFs (*orf297*, *orf296*, *orf194*, *orf138*, and *orf51*) were predicted to encode one or two transmembrane helices (Supplementary Fig. S1). *orf296* was immediately downstream from a repeat (929,551–929,873) that overlapped with the *atp1* gene, thus, the first 27 bp of *orf296* and *atp1* was identical (Supplementary Fig. S2). Observation of the corrected long reads supports a putative rearrangement at the repeats (Supplementary Fig. S2).

In addition, we generated raw paired-end (PE) Illumina reads that enabled *de novo* assembly of the complete plastome and draft mitogenome for the four *Hepatica* species and two related species *A. narcissiflora* and *P. koreana* from tribe Anemoneae. The additional *Hepatica* plastome sizes ranged from 159.8 to 160.9 kb, and the organization and gene/intron content were consistent with the results for *H. maxima* (Supplementary Table S4). The plastomes of *A. narcissiflora* and *P. koreana* showed minimal variation in genome size, IR boundary shifts, and GC content in comparison to those in the genus *Hepatica* (Supplementary Table S4). However, the plastome of *A. narcissiflora* lacked the *rps16* intron, and *P. koreana* contained an intact *rps16* gene in its plastome. For all seven taxa, 76 plastid-encoded genes were obtained from the assembled plastomes. Using BLAST searches between mitochondrial gene sequences from *H. maxima* and draft mitogenomes, 37 mitochondrion-encoded genes were obtained for all species (Supplementary Table S5).

**Genomic changes in plastid and mitochondrial genome evolution.** A Mauve alignment between the tribe Anemoneae and the outgroup (*Ranunculus macranthus*) identified six locally collinear blocks (LCBs) with five breakpoints that occurred at *rps16-trnQ*, *trnS-trnG*, *trnG-rps4*, *rps4-trnT*, and *ndhC-trnV* (Supplementary Fig. S3). An inferred plastome rearrangement model suggests that tribe Anemoneae has experienced six inversions (Fig. 3). The four inversion events (IV1 to IV4) occurred in the common ancestor of tribe Anemoneae, and two inversion events (IV5 and IV6) are specific to *Clematis*. The IR expansion is a synapomorphic event in tribe Anemoneae. A comparative analysis of the gene and intron contents revealed that losses of the



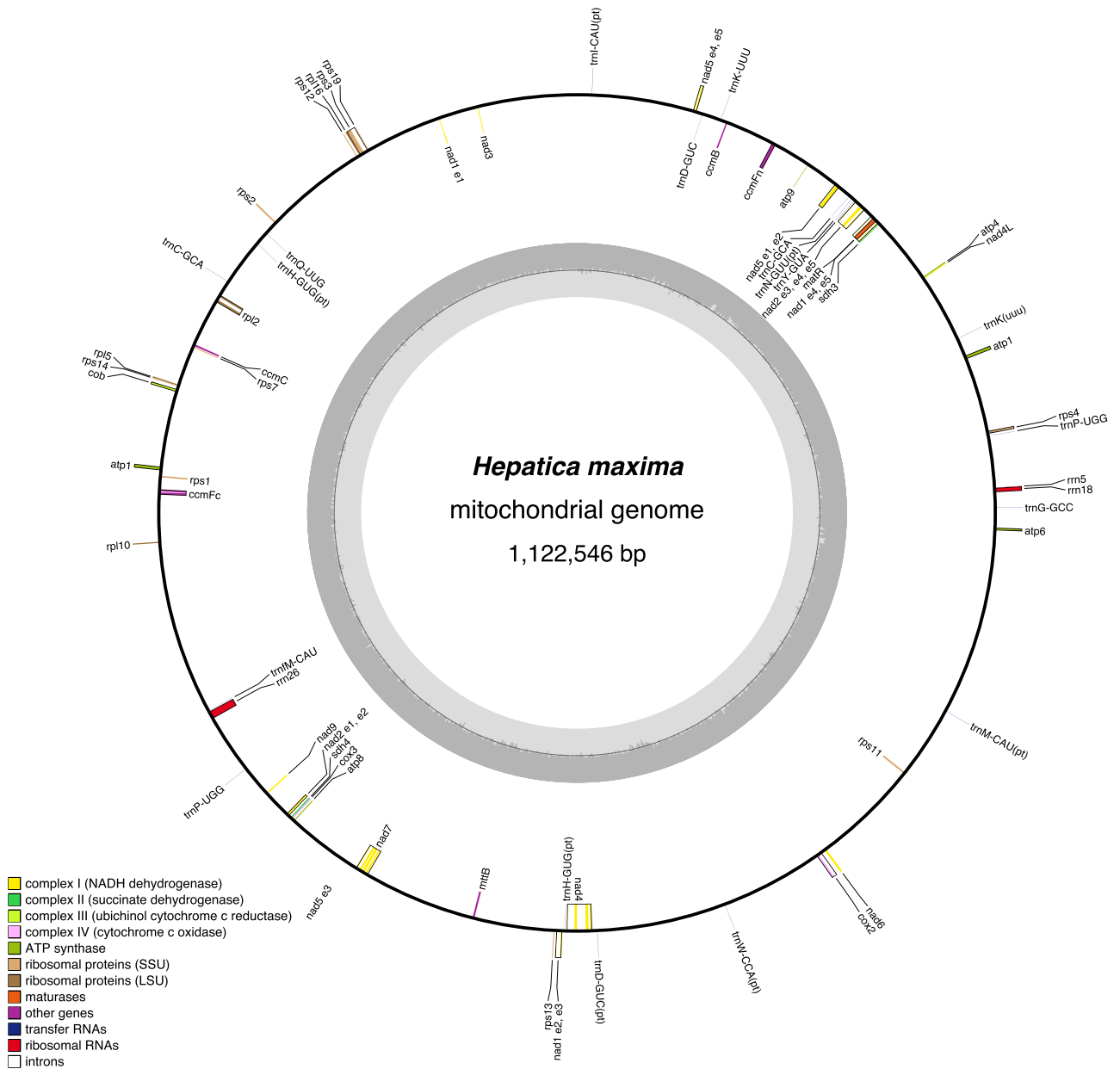
**Figure 1.** Map of the plastid genome of *Hepatica maxima*. Genes on the inside and outside of each map are transcribed in the clockwise and counterclockwise directions, respectively. The thick lines on the plastid map indicate the inverted repeats (IRA and IRb) that separate the genome into large and small single-copy regions. Ψ denotes a pseudogene. Black arrows indicate expansion events. Colored arcs on the outside of the map correspond to the locally collinear blocks inferred by Mauve (see Supplementary Fig. S3). The figures were constructed in OGDRAW v1.3.1 (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) and InkScape v0.92.2 (<https://inkscape.org>).

plastid *infA*, *rpl32* and *trnT-UGU* genes are shared by all analyzed species in the tribe Anemoneae, whereas the loss of *rps16* is unique to the *Hepatica* plastome (Fig. 3).

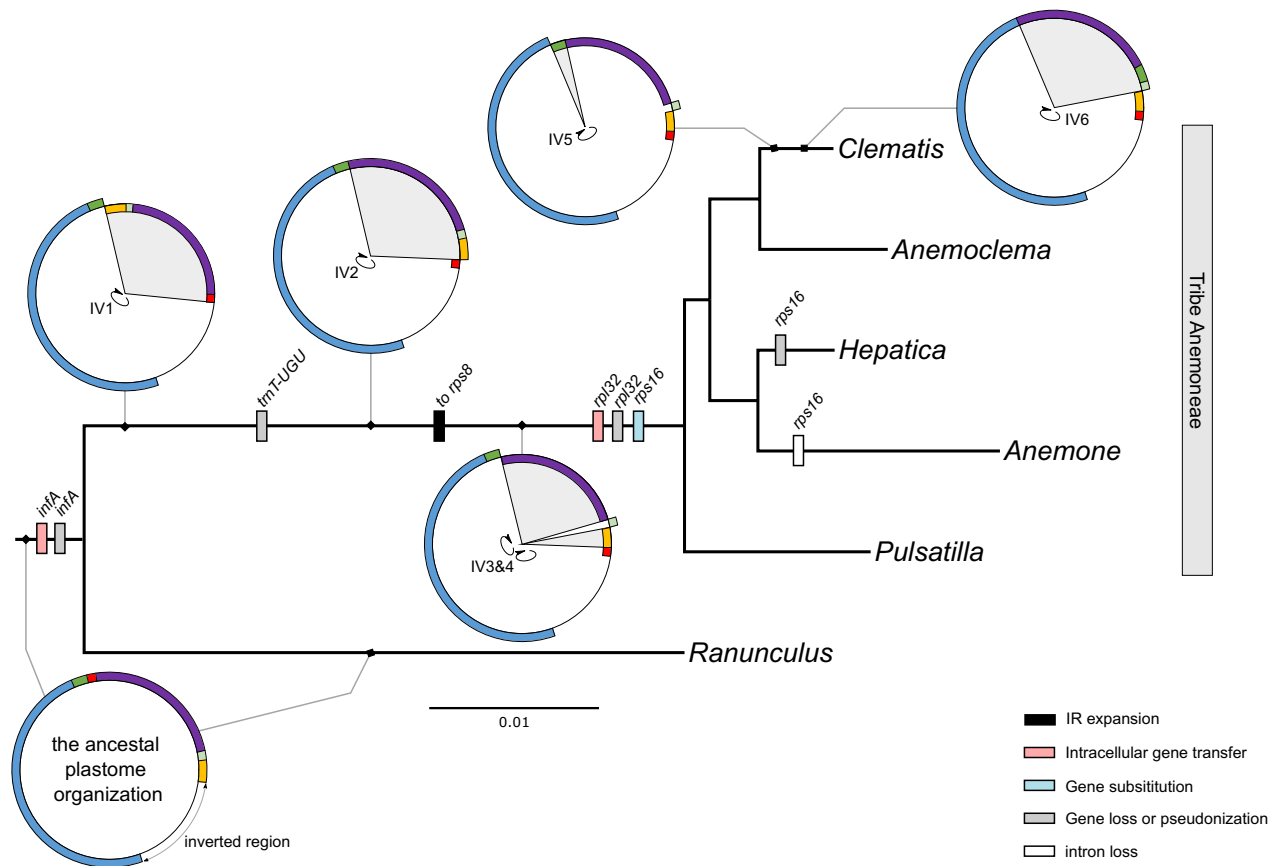
We identified putative losses of three plastid (*infA*, *rpl32*, and *rps16*) and two mitochondrial (*rps10* and *rps14*) genes in the *H. maxima* organellar genomes. To examine functional transfers to the nucleus, blastn searches against the *H. maxima* transcriptome with the five organellar genes of related species were performed. All predicted ORFs of the transcripts acquired novel targeting sequences and conserved domains (Fig. 4 and Supplementary Table S6). Six additional transcriptomes provided strong evidence for the functional transfer of five organellar genes into the nucleus within tribe Anemoneae (Supplementary Table S6). In particular, we found a transcript for the nuclear-encoded plastid *rps16* gene in all additional transcriptomes, although the plastid-encoded *rps16* gene has been lost in the *Hepatica* lineage (Supplementary Table S6).

	Plastome	Mitogenome
Genome size (bp)	160,876	1,122,546
Protein coding genes	76	39
rRNA genes	4	3
tRNA genes	29	17 (6)
Introns		
<i>cis</i> -spliced	19	18
<i>trans</i> -spliced	1	5
GC content (%)	37.8	46.2
Repeat content (%)	0.48	6.94
Coding content (%)	50.37	3.28

**Table 1.** Summary of *Hepatica maxima* plastid and mitochondrial genomes.



**Figure 2.** Map of the mitochondrial genome of *Hepatica maxima*. Genes on the inside and outside of each map are transcribed in the clockwise and counterclockwise directions, respectively. The figures were constructed in OGDRAW v1.3.1 (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>).

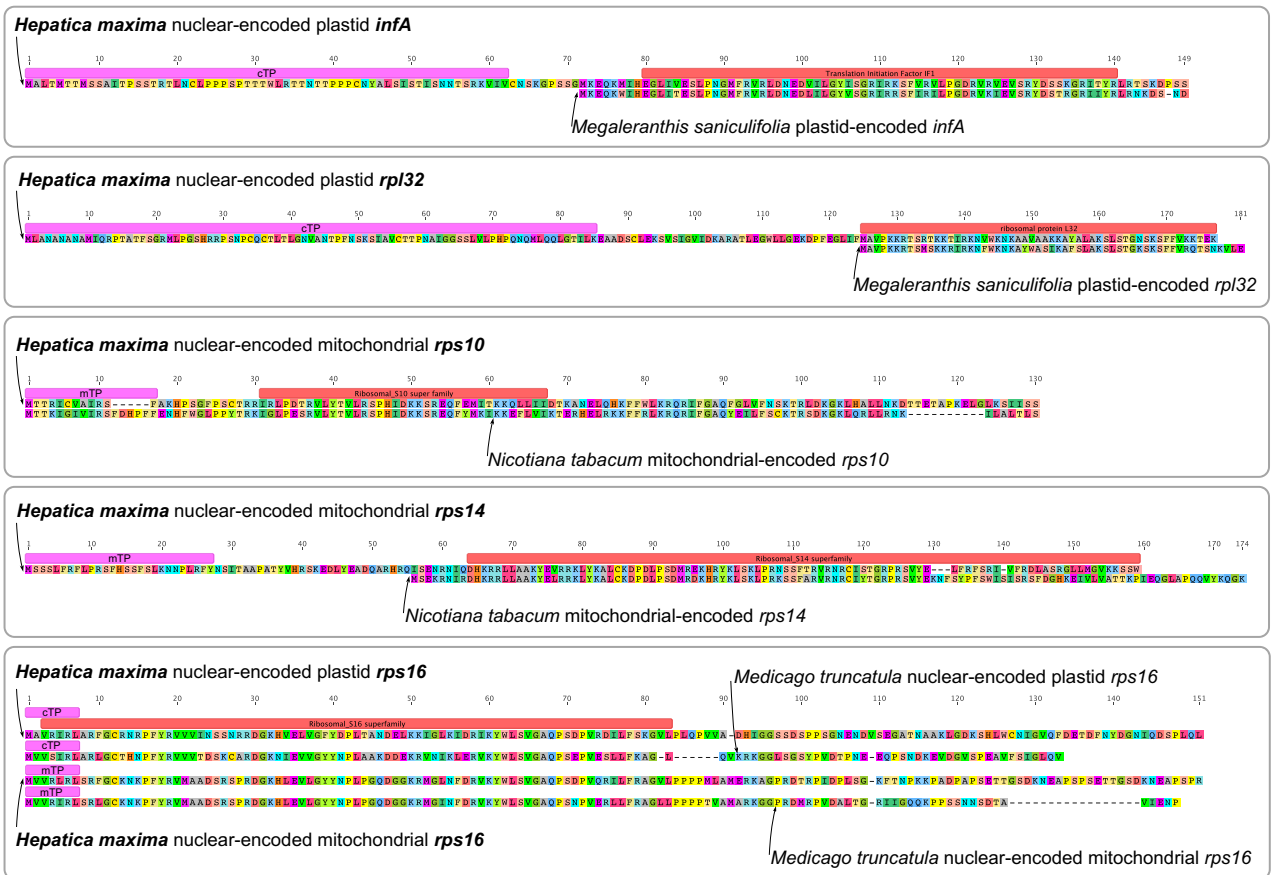


**Figure 3.** Summary of plastome rearrangement and intracellular gene transfer events in tribe Anemoneae with gene/intron loss events. The colors on the circular maps correspond to the locally collinear blocks inferred by Mauve (see Fig. 1 and Supplementary Fig. S3). IV; inversion. The figures were constructed in InkScape v0.92.2 (<https://inkscape.org>).

**Identification of nuclear orthologs.** To enable comparisons of the nuclear genes from *Hepatica*, we further sequenced the transcriptomes of the five *Hepatica* species and two outgroups, *A. narcissiflora* and *P. koreana*. We assembled the transcriptomes de novo using Trinity, excluded a transcript that included structural and base errors, and clustered transcripts with a high level of sequence similarity (Supplementary Table S7). We identified candidate coding regions within the clustered transcripts, and extracted complete ORFs (Supplementary Table S7). A total of 69,294 putative ORFs from seven species were assigned to 11,456 orthogroups by OrthoFinder (Supplementary Table S7). A total of 1428 orthogroups were shared across all seven analyzed species, and we obtained 510 single-copy orthologs for further analysis (Supplementary Table S7). Genes showing high degrees of misleading signals as indicated by the four parameters were filtered by TreSpEx (Supplementary Table S8), resulting in 413 single-copy orthologs.

**Phylogenomic reconstructions and divergence times.** All analyzed *Hepatica* species and the two related species *A. narcissiflora* and *P. koreana* shared 76 plastid-encoded, 37 mitochondrial-encoded, and 413 nuclear single-copy genes. To examine the origin of *H. maxima*, we performed phylogenomic analyses using two coalescent- and concatenation-based approaches.

The topologies of the Accurate Species TRee Algorithm (ASTRAL) and maximum likelihood (ML) supermatrix based on 526 genes were identical, with high support according to local posterior probability (LPP = 1.00) and bootstrap (BS = 100) values (Fig. 5). The short internal branches in the ML trees suggested rapid diversification in the *Hepatica* lineage (Supplementary Fig. S4). The analyses showed that *H. henryi* was basal to the rest of *Hepatica* with strong bootstrap support (LPP = 1.00 and BS = 100). *Hepatica maxima* and *H. nobilis* var. *japonica* as well as *H. asiatica* and *H. insularis* grouped together with high support values. However, the placement of *H. maxima* varied among the nuclear, plastid and mitochondrial ASTRAL and concatenated ML trees (Fig. 5). Both nuclear analyses robustly supported *H. maxima* as sister to *H. nobilis* var. *japonica* (LPP = 1.00, BS = 100, and gCF = 70.4), although the two organellar datasets were not consistent with the nuclear phylogeny. For example, the plastid ASTRAL and concatenated trees were consistent and placed *H. maxima* sister to *H. asiatica* and *H. insularis* with low support values (LPP = 0.59, BS = 63, and gCF = 10.5, Fig. 5). The mitochondrial ASTRAL tree revealed *H. maxima* as sister to *H. asiatica*, whereas the mitochondrial concatenated tree showed *H. maxima* as sister to the remaining three *Hepatica* species with weak support values (LPP = 0.53 and BS = 46, Fig. 5). The plastid and mitochondrial concordance factors were low for the four species, suggesting a high degree of conflict



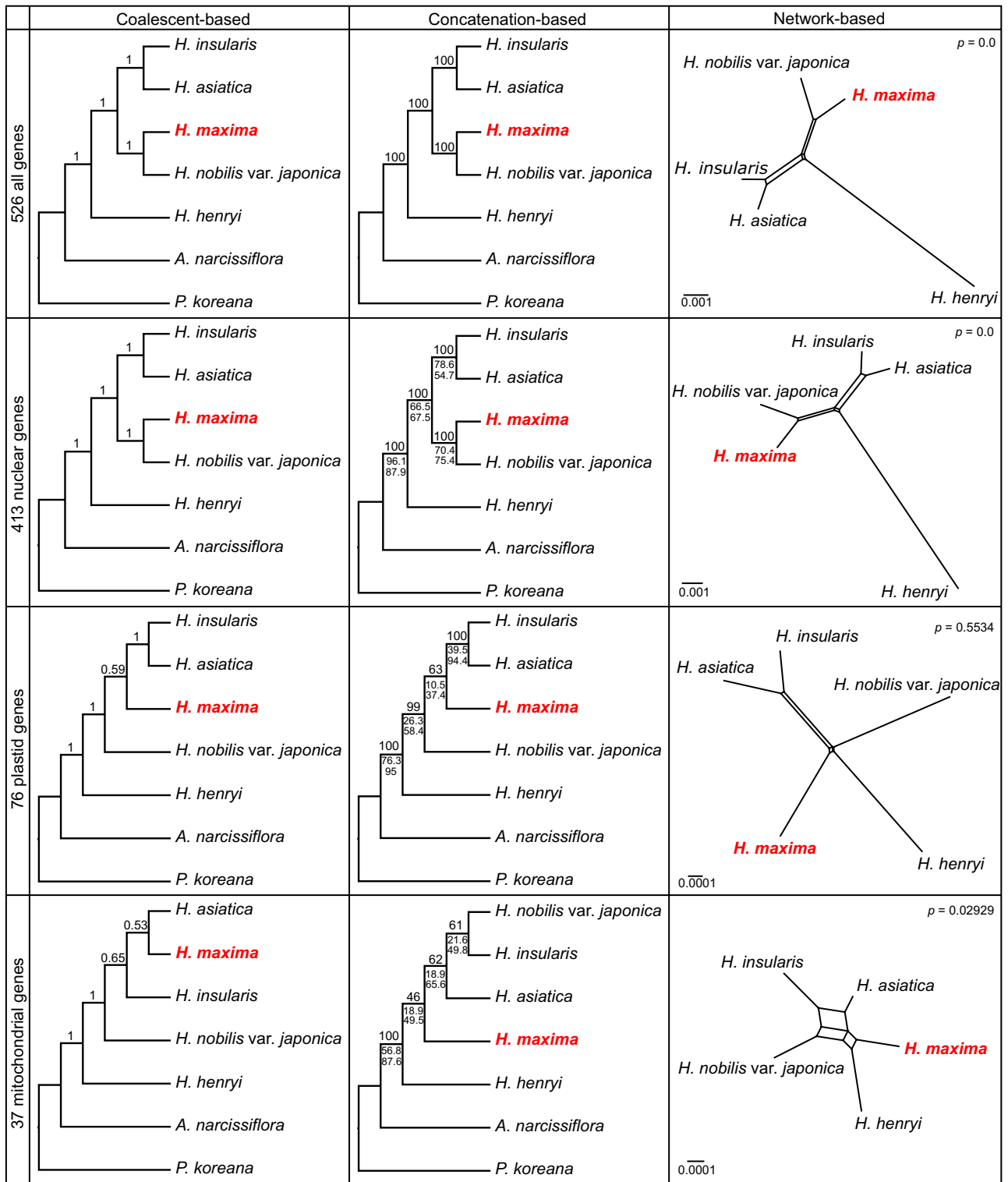
**Figure 4.** Amino acid sequence alignments of nuclear-encoded organellar genes in *Hepatica maxima* and their organellar copies in related species. Red boxes indicate the conserved domains, and pink boxes in the N-terminus indicate a transit peptide. The figures were constructed in Geneious R7 v7.1.8 (<https://www.geneious.com>).

among the gene trees (Fig. 5). Among the organellar gene trees, we only found five plastid (6.6% of 76 plastid genes) and two mitochondrial gene trees (5.4% of 37 mitochondrial genes) that were identical to the nuclear tree, which supported *H. maxima* as sister to *H. nobilis* var. *japonica*. All four split networks showed reticulate topologies, although the pairwise homoplasy index (PHI) test identified recombination signals in the supermatrix, nuclear and mitochondrial datasets (Fig. 5). The ML analysis of nucleotides after removing recombination sites by Gubbins supported the gene tree topology, which was identical to the ASTRAL and supermatrix phylogenies.

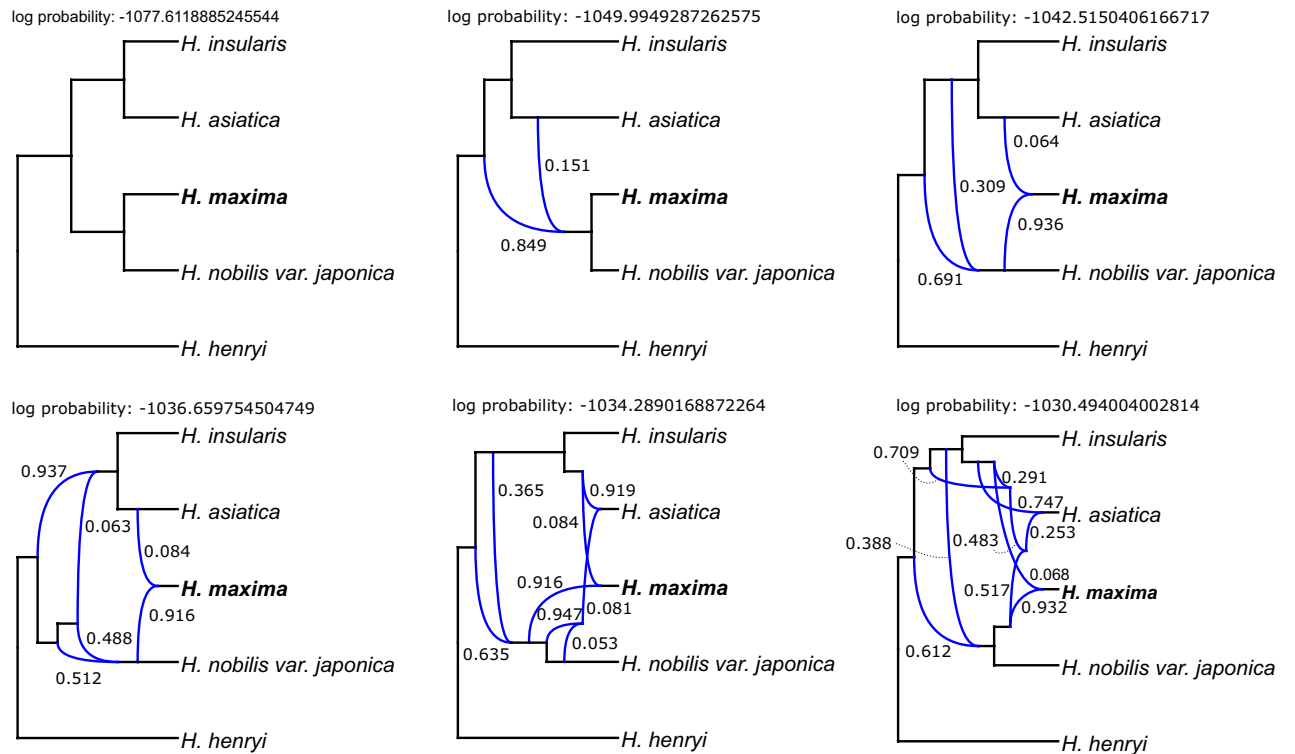
We further used the 526-gene dataset, excluding the two outgroups, to infer reticulate evolution in *Hepatica* by PhyloNet. PhyloNet analyses revealed a complex network of ancestral hybridization among the Asian lineages of *Hepatica*, with five reticulations selected by the Akaike information criterion as the best fit (Fig. 6 and Supplementary Table S9). The PhyloNet network with the highest likelihood (-1030.49400400281) indicated that ancient hybridization and introgression events occurred within four species, thus revealing a complicated evolutionary history of *H. asiatica* that involved hybridization and introgression among extinct species (Fig. 6). The gene flow events and divergence time estimates for *H. maxima* are summarized in Fig. 7. Our results indicated that the most recent common ancestor of crown *Hepatica* appeared ~2.03 Mya (95% highest posterior density (HPD) = 1.07–3.08, Supplementary Fig. S5). One of the lineages split into two new lineages 1.00 Mya (95% HPD = 0.64–1.45, Supplementary Fig. S5): (1) the ancestral lineage of *H. maxima*/*H. nobilis* var. *japonica* and the ancestral lineage of *H. insularis*/*H. asiatica*. The inheritance probability showed that the ancestral lineage of *H. maxima*/*H. nobilis* var. *japonica* had a genomic contribution of 38.8% from the ancestral lineage of *H. insularis*/*H. asiatica* between 0.34 and 1 Mya and that a minority of the genetic material composing the *H. maxima* lineage (6.8%) was derived from the ancestral branch of *H. asiatica*, which occurred <0.34 Mya (Fig. 7).

### Discussion

Oceanic islands provide good model systems for studying divergence and plant speciation because they are geologically young and isolated and have diverse habitats<sup>1</sup>. Many island plants have diverse and complex evolutionary histories involving various processes, such as rapid adaptation, hybridization, and introgression<sup>31–33</sup>. Genome-scale data provide a powerful source for studying evolution in island plants<sup>34</sup>. In this study, we first generated the complete plastid and mitochondrial genomes of *H. maxima*, one of the endemic plants observed on the oceanic Ulleung island, Korea, which is known to have originated via anagenetic speciation. Compared with



**Figure 5.** Phylogenomic trees of the nuclear, plastid, and mitochondrial genes inferred with ASTRAL-III, IQ-TREE2, and SplitsTree4. Local posterior probability and maximum likelihood bootstrap support values are shown above the branches on each cladogram. Numbers below the branches of maximum likelihood trees represent gene and site concordance factors, sequentially. The figures were constructed in ASTRAL v5.7.1 (<https://github.com/smirarab/ASTRAL>), IQ-TREE v2.0 (<https://www.iqtree.org/>), SplitsTree v4.15.1 (<https://www.splitstree.org/>) and modified in InkScape v0.92.2 (<https://inkscape.org>).



**Figure 6.** Phylogenetic networks inferred by PhyloNet using maximum likelihood under zero to six reticulation models. Blue branches indicate lineages involved in reticulated histories, and numerical values are the inheritance probabilities for each reticulation. The figures were constructed in PhyloNet (<https://bioinfocsc.rice.edu/phyloNet>) and modified in InkScape v0.92.2 (<https://inkscape.org>).



**Figure 7.** Summary of two introgression events associated with the origin of *H. maxima* on chronogram. Numbers at nodes indicated divergence time estimate in Mya. See supplementary Fig. S5 for more detail for tribe Anemoneae. Arrows indicate introgressions with the inferred inheritance probability. Color areas indicate the geographic distribution of each species. The images were constructed in InkScape v0.92.2 (<https://inkscape.org>) based on the chronogram. The top right corner of the photos shows habit, flower, and fruits (mature green) of *H. maxima*, and the bottom right corner of the photos shows the habit and hermaphroditic and female flowers of *H. asiatica*.



the ancestral angiosperm genome, the *H. maxima* plastome exhibits structural divergence, although its organization is identical to that previously described in *Hepatica* species<sup>35</sup>. Additional genome sequences from *Hepatica* revealed that plastomes of the *Hepatica* lineage have low rates of sequence variation and structural evolution. A comparison of the *Hepatica* plastomes with those of four representative genera from tribe Anemoneae provides insight into structural evolutionary events, including inversions and gene losses (Fig. 3). Plastome rearrangements in tribe Anemoneae have been documented<sup>35,36</sup>. Our results are consistent with the previous inference of six inversion events<sup>36</sup>, although the order of inversion events is slightly different. Our results suggested that the loss of *trnT-UGU* may be associated with one of these inversion events. The losses of the plastid *infA* and *rpl32* genes in tribe Anemoneae were revealed as the result of ongoing functional transfer into the nucleus. The finding of the nuclear-encoded plastid *rps16* gene was found in all additional transcriptomes, suggesting that functional replacement by gene substitution occurred in the common ancestor of tribe Anemoneae (Fig. 3). The coexistence of plastid and nuclear *rps16* homologs within tribe Anemoneae (except in *Hepatica*) indicates that functional transfer to the nucleus is necessary before the loss of the original plastid copy.

In angiosperms, the mitogenomes of approximately one hundred genera have been sequenced, which only represents 0.77% of 13,164 genera<sup>37</sup>, thus increasing the difficulty of performing comparative analyses within species, genera, and families. The *H. maxima* mitogenome represents the first complete mitogenome from the order Ranunculales and shows a relatively large size compared to that of the *Nelumbo nucifera* mitogenome (524,797 bp, NC\_030753) from Proteales and exhibits the loss of two mitochondrial genes. We also found evidence of mitochondrion-to-nucleus transfer of *rps10* and *rps14* in *H. maxima* (Fig. 4). The *H. maxima* mitogenome contains eighteen chimeric ORFs, five of which have partial mitochondrial genes and transmembrane helices. In particular, *orf297* and *orf296* include a portion of *atp1*, suggesting that these ORFs have the potential to cause cytoplasmic male sterility (CMS)<sup>38</sup>. The chimeric ORFs with an ATP synthase subunit are frequently associated with genome rearrangements<sup>38,39</sup>. In particular, *orf296* overlapped with one copy of repeats that are collapsed to call the other contigs (Supplementary Fig. S2), which may indicate that the repeats are associated with alternative arrangements resulting from homologous recombination. The chimeric structure of a partial *atp1* and association with the repeat are similar to that of known CMS-associated ORFs in *Mimulus*<sup>39</sup>. Further work is needed to determine if these ORFs are expressed and functional. The finding of a candidate CMS gene in the *H. maxima* mitogenome is of particular interest due to the presence of gynodioecy within the genus, which has been documented in *H. nobilis*<sup>40</sup>. We also observed both hermaphroditic and female individuals within a natural population of *H. asiatica* (Fig. 7) in southern South Korea. The candidate CMS gene may be useful for studies of gynodioecious populations of *H. asiatica*.

Large-scale genomic data enabled us to understand the evolutionary history of Asian *Hepatica*, including the insular endemic species *H. maxima*. Our results showed clear evidence of cytonuclear and organellar discordance (Fig. 5), indicating that Asian *Hepatica* may have experienced complicated evolutionary events. For example, the conflict between two organellar genomes suggests differential inheritance of organelles in the Asian *Hepatica* lineages, which resulted in the disruption of linkage disequilibrium (LD) between plastids and mitochondria<sup>41</sup>. The presence of CMS in gynodioecious systems can select for the paternal leakage of their mitochondrial genomes<sup>21</sup>. In Asian *Hepatica*, a gynodioecious population and potential CMS genes were observed, supporting the possibility of mitochondrial heteroplasmy in the lineages. Consequently, patterns of cytonuclear discordance in the Asian *Hepatica* lineages were also identified (Fig. 5). Cytonuclear discordance is known to be caused by hybridization and introgression, and these phenomena are common in plants because of organellar genome capture<sup>42</sup>. Taken together, the phylogenomic incongruence and the potential CMS implied that Asian *Hepatica* has experienced reticulation events (such as hybridization, introgression and horizontal gene transfer) or stochastic processes, including incomplete lineage sorting (ILS), via the biparental inheritance of organelles, although the incongruence is likely caused by insufficient parsimony-informative sites from the plastid and mitochondrial genes. The rapid diversification and gene tree discordance also indicated reticulate evolution in the Asian *Hepatica* lineage.

Our PhyloNet results provide strong evidence that hybridization and introgression events have contributed to the evolutionary history of Asian *Hepatica* (Fig. 6). In particular, *H. asiatica* has an extremely complex history (Fig. 6). The species network revealed that the evolution of the insular endemic species *H. maxima* involved two events: (1) massive ancient introgression (38.8%) from the common ancestor of *H. asiatica*/*H. insularis* to the common ancestor of *H. maxima*/*H. nobilis* var. *japonica* and (2) recent introgression (6.8%) from *H. asiatica* to *H. maxima* (Fig. 7). Recent introgression suggested that the ancestral population of *H. maxima* would have been very morphologically similar to the ancestor of *H. asiatica* and that interbreeding was possible. Genetic changes due to ancient and recent introgression events may be an important source for anagenetic speciation on Ulleungdo. During anagenetic speciation, *H. maxima* developed increased leaf, bract and seed sizes compared to those of the putative progenitor, which is a process known as “insular gigantism”. Functional analyses of polymorphisms in the *H. maxima* transcriptome associated with cell division (e.g., leaf, bract and seed sizes) require further study to provide valuable insights into adaptation to islands. A previous study suggested that *H. maxima* was derived from the Korean mainland species *H. asiatica*<sup>12</sup>. However, our results suggested that the insular endemic species *H. maxima* originated from an ancestral lineage that split from the common ancestor of *H. maxima*/*H. nobilis* var. *japonica*.

## Materials and methods

**DNA and RNA sequencing.** Hermaphroditic *Hepatica maxima* individuals were collected from Ulleungdo, Korea. We prepared total genomic DNA (gDNA) from fresh leaf tissue using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). The DNA was sequenced from one PE read from a 550 bp library and two mate pair (MP) reads from 3000 and 8000 bp libraries using the Illumina HiSeq 2000 sequencing platform (Illumina, San Diego, CA). Approximately 16 Gb of PE, 17 Gb of 3 kbp MP, and 10 Gb of 8 kbp MP reads were generated.

Additional gDNA was extracted using a Bead Genomic DNA Prep Kit For Plants (A\_type, BIOFACT Co., Daejeon, South Korea), and long reads were sequenced using the ONT GridION platform (ONT, Oxford, United Kingdom). Total RNA was isolated as described by Zhang et al.<sup>43</sup> and sequenced using the Illumina HiSeq 2000 sequencing platform.

Additional gDNA and RNA extractions from four *Hepatica* species (*H. asiatica*, *H. henryi*, *H. insularis*, and *H. nobilis* var. *japonica*) and the two outgroups *A. narcissiflora* and *P. koreana* were performed as described above. The six sets of gDNA and RNA were sequenced in only the PE library using the Illumina platform.

**Organellar genome assembly and annotation.** To complete the seven plastomes, the PE reads from gDNA were assembled de novo with Velvet v1.2.10<sup>44</sup> using multiple *k*-mers (69 to 95). For the *H. maxima* mitogenome, we performed hybrid assemblies with the PE/MP and Nanopore reads using Spades v3.13.1<sup>45</sup> and MaSURCA v3.2.6<sup>46</sup>. The contigs from the multiple assemblies were aligned manually in Geneious R7 v7.1.8 ([www.geneious.com](http://www.geneious.com))<sup>47</sup>, and the mitogenome was completed by tracking and end inspecting the contigs. To assess the depth of coverage, the Illumina and Nanopore reads were mapped to the genomes with Bowtie2 v2.2.9<sup>48</sup> and Minimap2 v2.17<sup>49</sup>, respectively. Plastome annotation was performed as described previously<sup>50</sup>, and the mitogenome was annotated using Mitofy<sup>51</sup>. Repeats were identified by performing BLAST + v2.10.0<sup>52</sup> comparisons of the *H. maxima* mitogenome against itself with an *e*-value cutoff of  $1 \times 10^{-6}$  and a word size of 16. Plastid-derived sequences in the mitogenome were identified by performing “blastn” searches of the *H. maxima* plastome against the mitochondrial genome with an *e*-value cutoff of  $1 \times 10^{-6}$ , a sequence identity of at least 80% and a minimum length of 50 bp. RNA editing sites were predicted using PREP-Mt<sup>53</sup> with a cutoff value of 0.5. To identify chimeric ORFs in the *H. maxima* mitogenome, all ORFs at least 150 bp in length were extracted from intergenic regions. We searched for ORFs containing one or more mitochondrial gene fragments using “blastn” with an *e*-value cutoff of  $1 \times 10^{-3}$ , a minimum length of 30 bp, and a sequence identity of at least 90%. Transmembrane domains in the identified chimeric ORFs were predicted using TMHMM v2.0<sup>54</sup>. The *H. maxima* organellar genome maps were drawn with OGDRAW v1.3.1<sup>55</sup>, and the genome sequences were deposited in GenBank under accession numbers MG952899 and MT568500.

For the other species, mitochondrial genes were identified in each draft assembly by a blastn search using the protein-coding genes from the *H. maxima* mitogenome as query sequences.

**Nuclear transcriptome assembly and orthology inference.** Five *Hepatica* and two outgroup de novo transcriptome assemblies were performed using Trinity v2.5.1<sup>56</sup> with default parameters. We then used TransRate<sup>57</sup> to filter “bad transcripts” that included chimeras, structural errors, incomplete assemblies, and base errors. To greatly reduce the number of redundant transcripts, highly similar copies among the filtered transcripts were clustered using CD-HIT v4.8.1<sup>58</sup> with an amino acid identity threshold of 0.95 (cd-hit-est). Finally, the clustered transcripts were filtered to obtain the longest isoform per Trinity gene. TransDecoder v5.5.0 (<https://transdecoder.github.io>) was then used to identify the best ORF for each transcript. Single-copy orthologs were inferred with OrthoFinder2 v2.3.11<sup>59</sup> using the multiple sequence alignment (MSA) program with MAFFT v7.271<sup>60</sup> and IQ-TREE v2.0<sup>61</sup>.

**Detection of functional replacement events.** IGT events were identified using “blastn” (*e*-value cutoff of  $1 \times 10^{-6}$ ) searches of the plastid-encoded genes (*infA* and *rpl32* from *Megaleranthis saniculifolia*; NC\_012615), mitochondrial-encoded genes (*rps10* and *rps14* from *Nicotiana tabacum*; NC\_006581), and nuclear-encoded genes (*rps16* from *Medicago truncatula*; AB365526) against the five *Hepatica* and two outgroup transcriptomes. TargetP v1.1<sup>62</sup> was used to predict plastid or mitochondrial transit peptides. The NCBI Conserved Domain Database (CDD) v3.18<sup>63</sup> was used for functional domain annotation. Nucleotide and amino acid sequences of nuclear and mitochondrial genes were aligned with MUSCLE<sup>64</sup> in Geneious R7.

**Phylogenomics and divergence time analyses.** Seventy-six plastid-encoded and thirty-seven mitochondrial-encoded genes were shared by the seven species (Supplementary Table S5). For nuclear genes, we excluded single-copy orthologs with possible sequence biases and misleading signals due to long-branch attraction and saturation using TreSpEx v1.1<sup>65,66</sup>. The individual gene alignments were generated based on the back-translation approach with MAFFT.

Phylogenomic analyses were performed with concatenation- and coalescent-based methods. For the concatenation method, nucleotide sequences of 526, 413, 76 and 37 genes of the five *Hepatica* species and two outgroups were concatenated into four large matrices. ML trees were constructed by IQ-TREE under the best-fit substitution model from each of the 526, 413, 76 and 37 partitioning schemes with the ultrafast bootstrap algorithm (1,000 replicates). The gene and site concordance factors (gCF and sCF, respectively) were also computed by IQ-TREE. For the coalescent-based method, single-gene trees were constructed using IQ-TREE under the best substitution model by ModelFinder, and then the program ASTRAL v5.7.1<sup>67</sup> was used for the datasets of 526, 413, 76 and 37 genes. The split network was inferred with SplitsTree v4.15.1<sup>68</sup> using the Uncorrected\_P and NeighborNet methods and 1000 bootstrap replicates. The  $\Phi_w$  (PHI) test<sup>69</sup> for detecting recombination was computed using SplitsTree. Recombination detected by Gubbins<sup>70</sup> was removed, and the resulting alignments were then analyzed by IQ-TREE to generate a ML phylogeny.

To reconstruct and evaluate the reticulate evolutionary relationships among the five *Hepatica* species, species networks that modeled ILS and introgression using a ML approach were constructed in PhyloNet v3.8.2<sup>71</sup> with the command “InferNetwork\_ML” and using the individual ML gene trees. Network searches were performed using only nodes in the rooted ML gene trees that had a bootstrap support of at least 75%, which allowed for zero

to six reticulations, and with optimization of the branch lengths and inheritance probabilities of the returned species networks under pseudo-likelihood.

We estimated the divergence times by using Bayesian Markov chain Monte Carlo (MCMC) methods in BEAST2 v2.6.2<sup>72</sup> with the concatenated alignment dataset. A calibration point estimate for tribe Anemoneae<sup>36</sup> (mean = 22.5, SD = 5.35, range = 13.7–31.3) was used with a normal prior distribution as the root age constraint.

### Data availability

Complete plastid and mitochondrial genomes and nuclear-encoded gene sequences have been submitted to GenBank (accession numbers MG952899, MT568500, MT560532–MT560561). The phylogenomic datasets generated during the current study are available in the Dryad Digital Repository, [<https://doi.org/10.5061/dryad.s1rn8pk50>].

Received: 5 June 2020; Accepted: 14 September 2020

Published online: 01 October 2020

### References

1. Takayama, K. *et al.* Genetic consequences of cladogenetic vs anagenetic speciation in endemic plants of oceanic islands. *AoB Plants* **7**, plv102. <https://doi.org/10.1093/aobpla/plv102> (2015).
2. Stuessy, T. F. *et al.* Anagenetic evolution in island plants. *J. Biogeogr.* **33**, 1259–1265. <https://doi.org/10.1111/j.1365-2699.2006.01504.x> (2006).
3. Whittaker, R. J., Triantis, K. A. & Ladle, R. J. A general dynamic theory of oceanic island biogeography. *J. Biogeogr.* **35**, 977–994. <https://doi.org/10.1111/j.1365-2699.2008.01892.x> (2008).
4. Kim, Y. K. Petrology of Ulreung volcanic island, Korea Part 1 Geology. *J. Jpn. Ass. Min.* **80**, 128–135. <https://doi.org/10.2465/ganko.1941.80.128> (1985).
5. Yang, S. *et al.* A floristic study of Ulleungdo Island in Korea. *Korean J. Pl. Taxon* **45**, 192–212. <https://doi.org/10.11110/kjpt.2015.45.2.192> (2015).
6. Takayama, K., Sun, B.-Y. & Stuessy, T. F. Anagenetic speciation in Ullung Island, Korea: genetic diversity and structure in the island endemic species, *Acer takesimensis* (Sapindaceae). *J. Plant Res.* **126**, 323–333. <https://doi.org/10.1007/s10265-012-0529-z> (2013).
7. Takayama, K., Sun, B.-Y. & Stuessy, T. F. Genetic consequences of anagenetic speciation in *Acer okamotoanum* (Sapindaceae) on Ullung Island, Korea. *Ann. Bot.* **109**, 321–330. <https://doi.org/10.1093/aob/mcr280> (2012).
8. Yang, J., Pak, J.-H., Maki, M. & Kim, S.-C. Multiple origins and the population genetic structure of *Rubus takesimensis* (Rosaceae) on Ulleung Island: Implications for the genetic consequences of anagenetic speciation. *PLoS ONE* **14**, e0222707–e0222707. <https://doi.org/10.1371/journal.pone.0222707> (2019).
9. Pfosser, M. *et al.* Evolution of *Dystaenia takesimana* (Apiaceae), endemic to Ullung Island, Korea. *Plant Syst. Evol.* **256**, 159–170. <https://doi.org/10.1007/s00606-005-0374-9> (2005).
10. Tamura, M. In *Die Natürlichen Pflanzenfamilien* (ed. Hiepko, P.) (Ducker & Humboldt, Berlin, 1995).
11. Sun, B. & Stuessy, T. Preliminary observations on the evolution of endemic angiosperms of Ullung Island, Korea. In *Evolution and Speciation of Island Plants* (eds Stuessy, T. & Ono, M.) 181–202 (Cambridge University Press, Cambridge, 1998).
12. Pfosser, M. *et al.* Phylogeny of *Hepatica* (Ranunculaceae) and origin of *Hepatica maxima* Nakai endemic to Ullung Island, Korea. *Stapfia* **95**, 16–27 (2011).
13. Nakai, T. Japanese *Hepatica*. *J. Jpn. Bot.* **13**, 308 (1937).
14. Jung, W.-C. & Heo, K. Fruit and seed morphology of Korean Ranunculaceae. *Korean J. Plant Taxon* **47**, 137–153. <https://doi.org/10.11110/kjpt.2017.47.2.137> (2017).
15. Knoop, V. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr. Genet.* **46**, 123–139. <https://doi.org/10.1007/s00294-004-0522-8> (2004).
16. Martin, W., Deusch, O., Stawski, N., Grünheit, N. & Goremykin, V. Chloroplast genome phylogenetics: why we need independent approaches to plant molecular evolution. *Trends Plant Sci.* **10**, 203–209. <https://doi.org/10.1016/j.tplants.2005.03.007> (2005).
17. Bock, D. G., Andrew, R. L. & Rieseberg, L. H. On the adaptive value of cytoplasmic genomes in plants. *Mol. Ecol.* **23**, 4899–4911. <https://doi.org/10.1111/mec.12920> (2014).
18. Birky, C. W. Jr. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci. USA* **92**, 11331–11338. <https://doi.org/10.1073/pnas.92.25.11331> (1995).
19. Mogensen, H. L. 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).
20. Thyssen, G., Svab, Z. & Maliga, P. Exceptional inheritance of plastids via pollen in *Nicotiana glauca* with no detectable paternal mitochondrial DNA in the progeny. *Plant J.* **72**, 84–88. <https://doi.org/10.1111/j.1365-3113.2012.05057.x> (2012).
21. McCauley, D. E. Paternal leakage, heteroplasmy, and the evolution of plant mitochondrial genomes. *New Phytol.* **200**, 966–977. <https://doi.org/10.1111/nph.12431> (2013).
22. Wolfe, K. H., Li, W. H. & Sharp, P. M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* **84**, 9054–9058 (1987).
23. Zimmer, E. A. & Wen, J. Using nuclear gene data for plant phylogenetics: progress and prospects. *Mol. Phylogenet. Evol.* **65**, 774–785. <https://doi.org/10.1016/j.ympev.2012.07.015> (2012).
24. Zhang, J., Xie, P., Lascoux, M., Meagher, T. R. & Liu, J. Rapidly evolving genes and stress adaptation of two desert poplars, *Populus euphratica* and *P. pruinosa*. *PLoS ONE* **8**, e66370. <https://doi.org/10.1371/journal.pone.0066370> (2013).
25. Chen, L. Y., Zhao, S. Y., Wang, Q. F. & Moody, M. L. Transcriptome sequencing of three *Ranunculus* species (Ranunculaceae) reveals candidate genes in adaptation from terrestrial to aquatic habitats. *Sci. Rep.* **5**, 10098. <https://doi.org/10.1038/srep10098> (2015).
26. Koenig, D. *et al.* Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. USA* **110**, E2655–2662. <https://doi.org/10.1073/pnas.1309606110> (2013).
27. Vigeland, M. D. *et al.* Evidence for adaptive evolution of low-temperature stress response genes in a Pooideae grass ancestor. *New Phytol.* **199**, 1060–1068. <https://doi.org/10.1111/nph.12337> (2013).
28. Metzker, M. L. Sequencing technologies: the next generation. *Nat. Rev. Genet.* **11**, 31–46. <https://doi.org/10.1038/nrg2626> (2010).
29. Treangen, T. J. & Salzberg, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* **13**, 36–46. <https://doi.org/10.1038/nrg3117> (2011).
30. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 239. <https://doi.org/10.1186/s13059-016-1103-0> (2016).
31. González-Pérez, M. A., Caujapé-Castells, J. & Sosa, P. A. Molecular evidence of hybridisation between the endemic *Phoenix canariensis* and the widespread *P. dactylifera* with random amplified polymorphic DNA (RAPD) markers. *Plant Syst. Evol.* **247**, 165–175 (2004).

32. Francisco-Ortega, J., Jansen, R. K. & Santos-Guerra, A. Chloroplast DNA evidence of colonization, adaptive radiation, and hybridization in the evolution of the Macaronesian flora. *Proc. Natl. Acad. Sci. USA* **93**, 4085–4090 (1996).
33. Friar, E., Cruse-Sanders, J. & McGlaughlin, M. Gene flow in *Dubautia arborea* and *D. ciliolata*: the roles of ecology and isolation by distance in maintaining species boundaries despite ongoing hybridization. *Mol. Ecol.* **16**, 4028–4038 (2007).
34. Crawford, D. J. & Archibald, J. K. Island floras as model systems for studies of plant speciation: prospects and challenges. *J. Syst. Evol.* **55**, 1–15. <https://doi.org/10.1111/jse.12234> (2017).
35. Liu, H. *et al.* Comparative analysis of complete chloroplast genomes of *Anemoclema*, *Anemone*, *Pulsatilla*, and *Hepatica* revealing structural variations among genera in tribe Anemoneae (Ranunculaceae). *Front. Plant Sci.* **9**, 1097. <https://doi.org/10.3389/fpls.2018.01097> (2018).
36. Zhai, W. *et al.* Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. *Mol. Phylogenet. Evol.* **135**, 12–21. <https://doi.org/10.1016/j.ympev.2019.02.024> (2019).
37. Christenhusz, M. J. M. & Byng, J. W. The number of known plants species in the world and its annual increase. *Phytotaxa* **261**, 17. <https://doi.org/10.11646/phytotaxa.261.3.1> (2016).
38. Hanson, M. R. & Bentolila, S. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *Plant Cell* **16**(Suppl), S154–169. <https://doi.org/10.1105/tpc.015966> (2004).
39. 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).
40. Kaul, M. L. *Male Sterility in Higher Plants* Vol. 10 (Springer Science & Business Media, Berlin, 2012).
41. Tsujimura, M., Mori, N., Yamagishi, H. & Terachi, T. A possible breakage of linkage disequilibrium between mitochondrial and chloroplast genomes during Emmer and Dinkel wheat evolution. *Genome* **56**, 187–193. <https://doi.org/10.1139/gen-2012-0153> (2013).
42. 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> (2017).
43. Zhang, J., Ruhlman, T. A., Mower, J. P. & Jansen, R. K. Comparative analyses of two Geraniaceae transcriptomes using next-generation sequencing. *BMC Plant Biol.* **13**, 228. <https://doi.org/10.1186/1471-2229-13-228> (2013).
44. 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).
45. 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> (2015).
46. 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).
47. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199> (2012).
48. 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).
49. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191> (2018).
50. Park, S., An, B. & Park, S. Recurrent gene duplication in the angiosperm tribe Delphinieae (Ranunculaceae) inferred from intracellular gene transfer events and heteroplasmic mutations in the plastid *matK* gene. *Sci. Rep.* **10**, 2720. <https://doi.org/10.1038/s41598-020-59547-6> (2020).
51. Alverson, A. J. *et al.* Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol. Biol. Evol.* **27**, 1436–1448. <https://doi.org/10.1093/molbev/msq029> (2010).
52. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421. <https://doi.org/10.1186/1471-2105-10-421> (2009).
53. Mower, J. P. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* **37**, W253–259. <https://doi.org/10.1093/nar/gkp337> (2009).
54. 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. <https://doi.org/10.1006/jmbi.2000.4315> (2001).
55. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.31: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **47**, W59–W64. <https://doi.org/10.1093/nar/gkz238> (2019).
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. Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M. & Kelly, S. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* **26**, 1134–1144. <https://doi.org/10.1101/gr.196469.115> (2016).
58. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565> (2012).
59. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238. <https://doi.org/10.1186/s13059-019-1832-y> (2019).
60. 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).
61. 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).
62. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**, 953–971. <https://doi.org/10.1038/nprot.2007.131> (2007).
63. Marchler-Bauer, A. *et al.* CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **41**, D348–352. <https://doi.org/10.1093/nar/gks1243> (2013).
64. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **5**, 113. <https://doi.org/10.1186/1471-2105-5-113> (2004).
65. Struck, T. H. TreSpEx-Detection of misleading signal in phylogenetic reconstructions based on tree information. *Evol. Bioinform. Online* **10**, 51–67. <https://doi.org/10.4137/ebo.s14239> (2014).
66. Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195. <https://doi.org/10.1093/molbev/mst024> (2013).
67. 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).
68. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267. <https://doi.org/10.1093/molbev/msj030> (2006).
69. Bruen, T. C., Philippe, H. & Bryant, D. A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665–2681. <https://doi.org/10.1534/genetics.105.048975> (2006).
70. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15. <https://doi.org/10.1093/nar/gku1196> (2015).

71. Wen, D., Yu, Y., Zhu, J. & Nakhleh, L. Inferring phylogenetic networks using PhyloNet. *Syst. Biol.* **67**, 735–740. <https://doi.org/10.1093/sysbio/syy015> (2018).
72. Bouckaert, R. *et al.* BEAST 2: A software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537. <https://doi.org/10.1371/journal.pcbi.1003537> (2014).

### Acknowledgements

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2015R1D1A3A01020009). We thank Dr. Kyoung Su Choi and Dr. Ku Tae Park for providing living materials of *Hepatica* species, and Dr. Sungwon Son for providing photographs of *H. maxima*.

### Author contributions

S.P. contributed to the design of the project and assembled, finished, and annotated the plastid and mitochondrial genomes, performed all analyses, prepared the figures and tables, and drafted the manuscript; and S.J.P. contributed to the design of the project and read/edited the manuscript. All authors read and approved the final draft of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-73397-2>.

**Correspondence** and requests for materials should be addressed to S.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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) 2020