Computational and Structural Biotechnology Journal 21 (2023) 5073–5091

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Extensive characterization of 28 complete chloroplast genomes of *Hydrangea* species: A perspective view of their organization and phylogenetic and evolutionary relationships

Gurusamy Raman [a], Kyoung-Su Choi [a,b], Eun Mi Lee [a], Clifford W. Morden [c], Hyeonah Shim [d], Jong-Soo Kang [d], Tae-Jin Yang [d,*], SeonJoo Park [a,*]

[a] Department of Life Sciences, Yeungnam University, Gyeongsan, Gyeongsan-buk, Republic of Korea
[b] Plant Research Team, Animal and Plant Research Department, Nakdonggang National Institute of Biological Resources, Sangju, Republic of Korea
[c] School of Life Sciences, University of Hawai'i at Mānoa, Honolulu, HI, USA
[d] Department of Agriculture, Forestry and Bioresources, Plant Genomics and Breeding Institute, Research Institute of Agriculture and Life Science, College of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Hydrangea
Biogeography
Disjunct distribution
Phylogenomics
Plastid genome

ABSTRACT

The tribe Hydrangeeae displays a unique, distinctive disjunct distribution encompassing East Asia, North America and Hawaii. Despite its complex trait variations and polyphyletic nature, comprehensive phylogenomic and biogeographical studies on this tribe have been lacking. To address this gap, we sequenced and characterized 28 plastomes of Hydrangeeae. Our study highlights the highly conserved nature of Hydrangeaceae chloroplast (cp) genomes in terms of gene content and arrangement. Notably, synapomorphic characteristics of tandem repeats in the conserved domain of *accD* were observed in the *Macrophyllae*, *Chinenses*, and *Dichroa* sections within the Hydrangeeae tribe. Additionally, we found lower expression of *accD* in these sections using structure prediction and quantitative real-time PCR analysis. Phylogenomic analyses revealed the subdivision of the Hydrangeeae tribe into two clades with robust support values. Consistent with polyphyletic relationships, sect. *Broussaisia* was identified as the basal group in the tribe Hydrangeeae. Our study also provides insights into the phylogenetic relationships of *Hydrangea petiolaris* in the Jeju and Ulleung Island populations, suggesting the need for further studies with more samples and molecular data. Divergence time estimation and biogeographical analyses suggested that the common ancestors of the tribe Hydrangeeae likely originated from North America and East Asia during the Paleocene period via the Bering Land Bridge, potentially facilitating migration within the tribe between these regions. In conclusion, this study enhances our understanding of the evolutionary history and biogeography of the tribe Hydrangeeae, shedding light on the dispersal patterns and origins of this intriguing plant group with its unique disjunct distribution.

1. Introduction

The family Hydrangeaceae, belonging to the order Cornales, encompasses two subfamilies, Jamesioideae and Hydrangeiodeae [29]. With over 270 accepted species, including well-known ornamental shrubs, Hydrangeaceae is distributed across Asia, North America, and southern Europe [66]. Previous studies have employed both morphological and molecular phylogenetic analyses to propose a revised classification for the genus *Hydrangea* within Hydrangeaceae [13,23,28,29,70]. These studies have illuminated the relationships among species and yielded valuable insights into the evolutionary history of this diverse plant family. According to the molecular phylogenetic analyses conducted using the *matK* marker [66], *Hydrangea* is polyphyletic to the eight groups of Hydrangeeae. More recent investigations, one utilizing 13 highly variable plastid markers (*trnK-trnH*, *trnK* gene and intron, *matK* gene, *rpS15-trnK* intergenic sequence (IGS), *rpS15* intron, *trnV-ndhC* IGS, *psbJ-petA* IGS, *psbT-petB* region, *petB-petD* region, *rpl16* intron, *rpl32-ndhF* IGS, *trnL-rpl33* IGS, and *ndhA* intron) [23] and the other utilizing one nuclear internal transcribed spacer (ITS) marker with four plastid markers [13], further supported the presence of two clades, referred to as *Hydrangea* I and *Hydrangea* II, within the tribe Hydrangeeae [13,23]. These analyses have enriched our comprehension of the

* Corresponding authors.
E-mail addresses: tjyang@snu.ac.kr (T.-J. Yang), sjpark01@ynu.ac.kr (S. Park).

https://doi.org/10.1016/j.csbj.2023.10.010
Received 5 September 2023; Received in revised form 4 October 2023; Accepted 6 October 2023
Available online 10 October 2023

evolutionary relations within the Hydrangeaceae family and have proposed a subdivision of the genus *Hydrangea* into 16 sections, including *Asperae*, *Broussaisia*, *Calyptranthe*, *Cardiandra*, *Chinenses*, *Cornidia*, *Cecumaria*, *Deinanthe*, *Dichroa*, *Heteromallae*, *Hirtae*, *Hydrangea*, *Macrophyllae*, *Pileostegia*, *Schizophragma*, and *Sylosae*. However, the position of *H. arborescens* [23] is not well supported and remains unresolved, and certain groups within the genus *Hydrangea* show polyphyly [13,23].

Recently, [13] proposed two possible strategies to address the polyphyletic nature of *Hydrangea*. The first option involves establishing new genera to accommodate these groups while keeping the eight satellite genera distinct. This approach would involve dividing *Hydrangea* into smaller units, potentially resulting in the creation of at least seven new genera, with two of them having only one species each. However, this would lead to the challenge of distinguishing morphologically similar taxa, especially within the split *Hydrangea* group. The level of division could vary depending on the acceptance of genera that have only one species or are paraphyletic. The second option involves incorporating the eight satellite genera into a broadly defined *Hydrangea*, resulting in a comprehensive monophyletic *Hydrangea* that includes all species. However, this approach would lead to a wide range of morphological variation within the newly formed *Hydrangea* genus, making it less practical to classify distinct-looking taxa as separate satellite genera. Additionally, this approach would require assigning new specific names to several taxa within the expanded *Hydrangea*. Additionally, [84] utilized 81 plastomes and extensively studied the phylogenomic position of the genus *Hydrangea* and revealed that some of the sections are paraphyletic.

Hydrangea petiolaris is a species within Hydrangeaceae that is restricted to Jeju and Ulleung Islands of Korea, Japan, and Sakhalin (Russia) [9]. [36] conducted a study that divided *H. petiolaris* into two distinct groups: one comprising individuals from Jeju Island (excluding Mt. Halla) and the other including individuals from Ulleung, Mt. Halla (Jeju), and Japan. These findings highlight the genetic differentiation within *H. petiolaris* populations across different geographical regions.

Over the past few years, there has been a growing utilization of cp genome sequences in the field of plant taxonomy to resolve phylogenetic implications [56,83,89]. The NCBI database now contains the published or deposited cp genomes of over ten *Hydrangea* species [18]. While cp genome analysis has provided insights into the phylogenomic relationships and their evolution within the order Cornales, including the genus *Hydrangea* [17,18], a comprehensive analysis of multiple *Hydrangea* species is still lacking.

Therefore, in this study, we aimed to contribute to the understanding of *Hydrangea* species by presenting the complete cp genomes of 28 individuals from 22 different Hydrangeeae species, two Philadelpheae species and one Jamesioideae species. Our objectives are as follows: (1) assess the utility of analyzed cp genome sequences for resolving phylogenomic relations between Hydrangeeae species with different datasets, such as the whole cp genome, protein-coding gene sequences (PCSs) and translated amino acid datasets, and a Bayesian tree; (2) compare the cp genomes of Hydrangeeae species and identify effective regions for phylogenetic analyses within the genus; (3) estimate the divergence times and biogeography of *Hydrangea* species; and (4) compare the *H. anomala* subsp. *petiolaris* populations of Jeju and Ulleung Islands of Korea. By undertaking these endeavors, our goal is to advance our comprehension of the evolutionary history and genetic distinctions within the Hydrangeaceae family, with a particular focus on the *Hydrangea* genus.

2. Materials and methods

2.1. Plant samples, sequencing and genome annotation

A total of twenty-eight individuals representing 25 species, including twenty-two *Hydrangea* species, two Philadelpheae species and one Jamesioideae species as the outgroup species, were selected for this study. Voucher specimens were deposited in the Yeungnam University Herbarium (YNUH), Gyeongsan, Republic of Korea. To investigate intraspecific variability within the plastid genome, we sequenced four *H. petiolaris* individuals, including three from Jeju Island and one from Ulleung Island. Comprehensive information on the sampled species in this study is shown in Supplementary Table S1.

Genomic DNA was extracted using the DNeasy Plant Mini Kit (Qiagen Inc., GmbH, Germany). The DNA samples were then sequenced on the HiSeq 2500 platform (Illumina, San Diego, CA, USA) at Phyzen Ltd., South Korea. The PE library (2 150 bp) was constructed using a TruSeq PCR-free kit, paired reads with a 550 bp insert size were sequenced, and ~18 GB of raw data were generated. The PE reads were utilized for *de novo* assembly using the Velvet v1.2.10 assembler [88]. Multiple k-mers were applied during the assembly process. To evaluate the depth of coverage, the Illumina reads were mapped to the respective genome using Bowtie v7.2.1 [39] in Geneious Prime v2022.1. Annotation of all plastomes was carried out using the online program Geseq [78] and Geneious Prime v2022.1. The entire genomes were visually mapped using the OGDRAW program [24].

2.2. Divergence hotspot identification

To examine the similarities and differences among the 40 *Hydrangea* chloroplast (cp) genomes, we performed an alignment using MAFFT software [34]. The mVISTA program was used to compare the complete 28 plastomes utilized in this study with an additional 12 *Hydrangea* species retrieved from the NCBI [16] using Shuffle-LAGAN mode. In this analysis, *H. arguta* was chosen as the reference genome. To assess nucleotide diversity (pi) and genetic variation within the Hydrangeeae and Hydrangeaceae cp genomes, we calculated the total number of mutations (eta) and the average number of nucleotide differences (K) in the protein-coding genes and the intergenic and intron-containing regions individually using DnaSP software v6 [64]. Gaps and missing data were excluded from this analysis.

2.3. Analysis of substitution rate

To analyze the synonymous (K_S) and nonsynonymous (K_A) substitution rates, we separately compared the cp genomes of 40 Hydrangeeae and 50 Hydrangeaceae species. Specifically, we extracted the exons of individual functional protein-coding genes from these genomes and performed separate alignments using Geneious Prime v2022.1. Following alignment, we translated the aligned sequences into protein sequences. Subsequently, we employed DnaSP v6 software [64] to assess the K_A and K_S substitution rates, excluding stop codons from the analysis.

2.4. Analysis of the accD gene

To evaluate the tandem repeats in the *accD* gene, all the *accD* genes from the Hydrangeaceae cp genomes were extracted, and multiple alignments were performed using the MAFFT alignment tool. On the basis of the multiple alignment analysis, species such as *H. paniculata*, *H. febrifuga* and *H. hirsuta* were selected for structure prediction and validation using real-time quantitative PCR (RT—qPCR).

To understand the structure of the accD protein, the *accD* gene was translated into a protein, and AlphaFold was utilized for 3D structure modeling [32]. The modeling process involved template selection, alignment score and model quality factors in a fully automated process. The obtained model was validated using tools such as PROCHECK, ProSA, ProQ, Molprobity and ERRAT, which provided reliable scores for structural compatibility [12,32,40,8,80,82].

The *accD* gene-specific primer was designed using Geneious Prime v2022.1 and synthesized by SFCprobes Co. Ltd., Cheongju-si, Chungcheongbuk-do, South Korea (Supplementary Table S2). RNA was extracted from the three selected species, and first-strand cDNA

synthesis was performed using the GoScriptTM Reverse Transcription System kit from Promega Corporation (Madison, WI, USA) according to the instruction manual. The obtained cDNA was utilized as the template for the qPCR analysis, which was performed with the GoTaq qPCR Master Mix from Promega Corporation (Madison, WI, USA). Three technical replicates were set up for each sample, and the total reaction system volume was 20 μl, consisting of 10 μl of GoTaq qPCR Master Mix (2 ×), 7 μl of nuclease-free water, 1 μl of each forward and reverse primer, and 1 μl of cDNA. The Applied Biosystems StepOnePlus Real-Time PCR System from Thermo Fisher Scientific (Waltham, MA, USA) was used for the qPCR analysis. The relative expression of the target gene in the samples was obtained with the $2^{-\Delta\Delta CT}$ calculation method. Statistical significance testing, such as t tests, was performed to analyze the obtained data and determine if there were significant differences in the expression levels of *accD* among the tested species.

2.5. Codon usage

The codon usage patterns were analyzed for all protein-coding genes in the cp genomes of Hydrangeaceae and Loasaceae. These patterns were visually represented as heatmaps for a total of 57 species from both plant families. Histograms were then created using the Heatmapper program, utilizing the relative synonymous codon usage (RSCU) values [3]. The RSCU values were determined by calculating the frequency of each codon relative to all codons encoding specific amino acids compared to their expected probabilities. A codon with an RSCU value below 1.0 indicated less frequent usage than expected, while a value above 1.0 indicated more frequent usage than expected.

2.6. Analysis of RNA editing

We utilized the online program PREPACT v2.0 for plants to identify potential RNA editing sites within the 57 protein-coding genes from 57 cp genomes [45]. For this analysis, *Nicotiana tabacum* served as the reference species, with the following parameters: a cutoff expect (E) value of 0.001, a minimum value of 8, and a filter threshold value of 70%.

2.7. Analysis of tandem repeats and single sequence repeats (SSRs)

To identify repeat sequences in the cp genomes of *Hydrangea*, we utilized REPuter software (Kurtz et al., 2001) with the following parameters: a Hamming distance of 3, a minimum sequence identity of 90%, and a requirement for repeat sequences to exceed 30 base pairs in length. Additionally, we employed the MISA tool (Beier et al., 2017) to detect simple sequence repeats (SSRs) within the 25 *Hydrangea* cp genomes. For SSR identification, the following criteria were applied: a minimum sequence length of 10 bp for mononucleotide repeats; five repeat units for dinucleotide repeats; four repeat units for trinucleotide repeats; and three repeat units for tetra-, penta-, and hexanucleotide repeats.

2.8. Phylogenetic analysis

We performed phylogenetic reconstruction using both maximum likelihood (ML) and Bayesian inference (BI) methods according to the concatenation of 79 protein-coding gene sequences from 57 taxa. Among the 57 taxa, 50 species were from Hydrangeaceae, and seven species from Loasaceae were used as the outgroup. The analysis utilized various datasets, including the whole cp genome, protein-coding gene sequences (PCSs) and translated amino acid datasets. RAxML software with GTRI-based nucleotide substitution was employed to analyze all nucleotide datasets for the ML analysis [72]. To provide statistical support for the branches, we conducted rapid analysis with 1000 bootstrap replicates. Additionally, BI analysis was carried out using MrBayes v3.2.6, employing 79 concatenated PCSs, the gamma model for rate variation, and the HKY85 substitution model [63].

2.9. Estimation of evolutionary rate

To analyze the divergence times of the Hydrangeaceae lineages, a BI approach with MCMC sampling was performed using BEAST v2.5 with some modifications [6]. In the analysis, a relaxed-clock log-normal model was applied. The Markov chain Monte Carlo (MCMC) chain comprised 200 million steps, with sampling conducted every 1000 generations, and a 10% burn-in phase was implemented. A GTR nucleotide substitution model was employed, featuring a gamma distribution with four rate categories. Divergence times and credibility intervals were estimated using a Yule prior. The assessment of the sample size was performed using Tracer v1.6 analysis software from the Institute of Evolutionary Biology at the University of Edinburgh, UK [61].

For the divergence time analysis, two calibration points were established. The first calibration point, marking the divergence of the Hydrangeaceae lineage, was set at 59.2 million years ago (Ma), with a 95% confidence interval ranging from 39.8 to 78.8 Ma. The second calibration point, indicating the divergence of the Loasaceae lineage, was established at 70.2 Ma, with a 95% confidence interval spanning from 52.6 to 86.7 Ma, as reported by [18]. Both calibrations were incorporated using a log-normal distribution. After the MCMC analysis, a maximum clade credibility (MCC) tree was generated using TreeAnnotator v2.1.2 [6].

2.10. Phylogenetic informativeness analysis

To assess the phylogenetic informativeness (PI) of each protein-coding gene, we employed the online program PhyDesign [48]. For DNA sequences, we utilized the HyPhy substitution rates algorithm [58] with default settings. To mitigate phylogenetic noise and accommodate gene length disparities, we applied the per-site profile approach recommended by [48]. We constructed phylogenetic trees using the maximum likelihood (ML) method, following [72], according to concatenated alignments of 79 protein-coding sequences obtained from 57 species of Hydrangeaceae and closely related species. Before inputting the ML trees into PhyDesign, we transformed them into rooted ultrametric trees using the "chronos" function from the ape package [57] implemented in R v.4.2.2. This involved calibrating the trees using an arbitrary time scale, with tips assigned a time of 0 and the root assigned a time of 1. Subsequently, the relative-time ultrametric tree and the alignment of 79 concatenated protein-coding genes served as the input files for PhyDesign to calculate phylogenetic informativeness values (PIVs).

2.11. Reconstructing historical biogeography

Biogeographic data for Hydrangeaceae and other Loasaceae outgroup species were obtained from the Plants of the World Online (POWO) database, maintained by the Royal Botanic Gardens, Kew, UK (https://powo.science.kew.org). The distribution range of Hydrangeaceae species was divided into five areas: Asia [A], South Africa [B], North America [C], Central America [D] and Oceania [E]. To conduct a more detailed analysis, Asia [A] was further subdivided into eastern Asia (A), southern Asia (B), southeastern Asia (C) and the Asian part of Russia (D). Additionally, North America [C] was divided into the subregions southeastern North America (F), southwestern North America (G), northeastern North America (H), western North America (I) and Canada (J), and Central America was divided into Mexico and Guatemala (K) and Peru and Chile (L), while South Africa (E) and Hawaii (M) were considered separate areas.

To analyze ancestral areas and evaluate the spatial patterns of geographic diversification within Hydrangeaceae, we utilized a range of analytical methods. These methods included statistical dispersal-vicariance analysis (S-DIVA), the Bayesian binary method (BBM),

dispersal-extinction-cladogenesis (DEC), the recently modified statistic DEC (S-DEC), and BayArea analysis, all implemented within the Reconstruct Ancestral State in Phylogenies (RASP) v4.2 software [62, 86]. Additionally, we utilized the BioGeoBEARS package to select the most appropriate model for our analysis. In these analyses, a dataset consisting of 79 plastid protein-coding genes was analyzed with BEAST software to generate a set of 100,000 trees from the MCMC output.

3. Results

3.1. General characteristics of the chloroplast genomes of 28 Hydrangeaceae species

In the present study, we sequenced a total of 28 individuals, including 25 individuals from 22 *Hydrangea* species and three outgroup species, to obtain their cp genomes using a *de novo* approach. The depth coverage of all the plastomes varied from 1186 × to 7806 × (Table 1). The complete cp genomes of the 28 species exhibited the typical circular, double-stranded structure and varied in length from 156,430 bp (*Kirengeshoma koreana*) to 157,974 bp (*H. viburnoides*) (Fig. 1). All 28

Table 1
General characteristics of 28 Hydrangeaceae plastomes.

Sl. No.	Species name	GenBank accession number	Depth coverage (×)	Total length (bp)	LSC (bp)	SSC (bp)	IR (bp)	GC content (%)	Total number of genes	Unique Genes (Duplicated)			
										Protein-coding	tRNA	rRNA	No. of introns
1	*Hydrangea alternifolia*	OR268908	2509	157806	86784	18770	26125	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
2	*Hydrangea arguta*	OR268909	3136	156827	85893	18708	26113	37.9	135	79 (8)	30 (8)	4 (4)	21 (5)
3	*Hydrangea obtusifolia*	OR268910	3300	157715	86819	18702	26097	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
4	*Hydrangea barbara*	OR268911	2363	157715	86819	18702	26097	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
5	*Hydrangea formosana*	OR268912	7806	157335	86398	18763	26087	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
6	*Hydrangea caerulea*	OR268913	2669	157721	86804	18757	26080	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
7	*Hydrangea damingshanensis*	OR268914	7254	157749	86828	18645	26138	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
8	*Hydrangea bifida*	OR268915	2576	157826	86883	18749	26097	37.7	135	79 (8)	30 (8)	4 (4)	21 (5)
9	*Hydrangea febrifuga*	OR268916	2741	157646	86727	18675	26122	37.9	135	79 (8)	30 (8)	4 (4)	21 (5)
10	*Deutzia paniculata*	OR268917	4933	157126	86372	18564	26095	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
11	*Hydrangea luteovenosa*	OR268918	5054	157494	86596	18646	26126	37.9	135	79 (8)	30 (8)	4 (4)	21 (5)
12	*Hydrangea hirsuta*	OR268919	2626	157652	86745	18644	26132	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
13	*Hydrangea arborescens*	OR268920	2171	157898	86898	18792	26104	37.9	135	79 (8)	30 (8)	4 (4)	21 (5)
14	*Hydrangea paniculata*	OR268921	4933	157823	86838	18739	26123	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
15	*Hydrangea petiolaris* jjch	OR268922	1896	157714	86722	18774	26109	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
16	jjaw	OR268923	2659	157714	86722	18774	26109	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
17	jjsr	OR268924	3170	157714	86722	18774	26109	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
18	ul	OR268926	1186	157693	86722	18753	26109	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
19	*Hydrangea quercifolia*	OR268925	4308	158163	86868	18393	26451	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
20	*Hydrangea scandens*	OR268927	3692	157525	86635	18638	26126	37.9	135	79 (8)	30 (8)	4 (4)	21 (5)
21	*Jamesia americana*	OR268928	3049	157882	87047	18691	26072	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
22	*Kirengeshoma koreana*	OR268929	3950	156430	86941	18563	25963	37.5	135	79 (8)	30 (8)	4 (4)	21 (5)
23	*Hydrangea viburnoides*	OR268930	1275	157974	85106	18702	27083	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
24	*Hydrangea platyarguta*	OR268931	2245	156716	85706	18734	26138	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
25	*Hydrangea schizomollis*	OR268932	1949	157630	86744	18688	26099	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
26	*Hydrangea hydrangeoides*	OR268933	1383	157554	86710	18706	26069	37.8	135	79 (8)	30 (8)	4 (4)	21 (5)
27	*Hydrangea macrophylla*	OL944390	4619	157617	86672	18697	26124	37.9	135	79 (8)	30 (8)	4 (4)	21 (5)
28	*Hydrangea serrata*	OL944391	2868	157743	86834	18701	26104	37.9	135	79 (8)	30 (8)	4 (4)	21 (5)

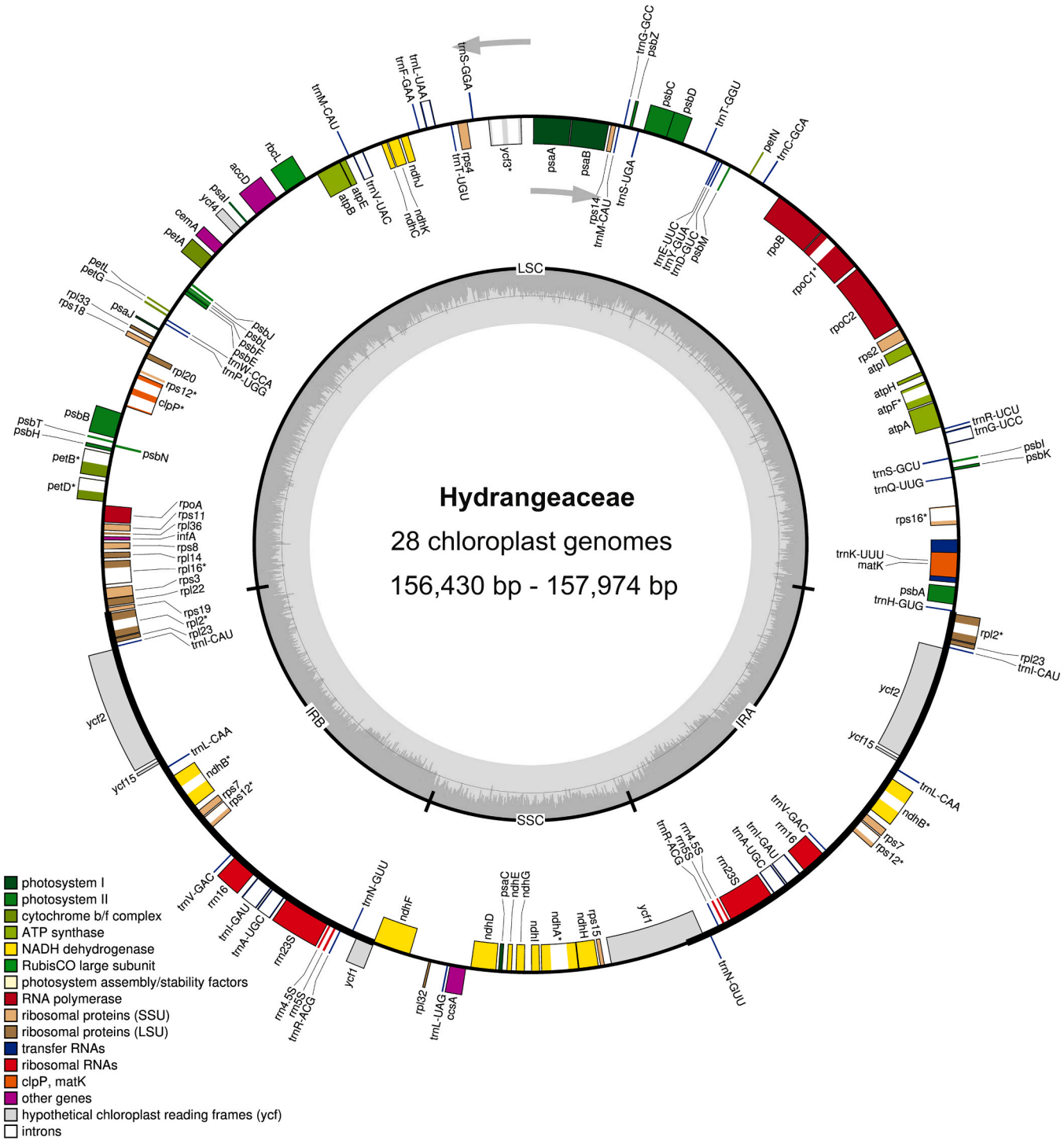

Fig. 1. Circular chloroplast genome map of 28 Hydrangeaceae species. In the diagram, distinct colors indicate genes with distinct functional groups. The genes in the circles are transcribed clockwise, and the genes outside the circles are transcribed counterclockwise. A large single copy (LSC) region, small single copy (SSC) region and two inverted repeat (IRa and IRb) regions are indicated in the inner circles. The light gray circles signify the A/T content, and the dark gray circles show the GC content.

sequences exhibited a quadripartite structure, which included a large single copy (LSC) region, a small single copy (SSC) region, and a pair of inverted repeat (IR) regions. The LSC region varied in length from 85,106 to 87,047 bp, with a GC content ranging from 35.6% to 36.1% (Table 1). The SSC region displayed a length distribution of 18,393–18,792 bp, with a GC content ranging from 31.2% to 31.8%. The IR region in the 28 Hydrangeaceae species ranged from 25,963 to 27,083 bp in length and exhibited a GC content of 42.8–43.2%. The complete cp genome sequences of these 28 Hydrangeaceae species have been deposited in GenBank under the accession numbers OR268908–OR268933 and OL944390–OL944391.

Despite variations in the lengths of the cp genomes among the 28 species, genetic composition analyses revealed commonalities. The gene positions are illustrated in Fig. 1. In total, 135 functional genes were predicted in all 28 Hydrangeaceae species, comprising 113 unique genes, including 79 protein-coding genes, 30 transfer RNA (tRNA)

genes, and four ribosomal RNA (rRNA) genes (Supplementary Fig. S1). These genes in the Hydrangeaceae species can be broadly categorized into photosynthesis-related genes, cp self-replication genes, and other genes (Supplementary Table S3). Of these genes, 18 intron-containing genes (ICGs) were identified, encompassing both protein-coding genes and six tRNA genes (Supplementary Fig. S1). Among the ICGs, *ycf3* and *clpP* possessed two introns each, while the remaining ICGs contained only one intron. The *rps11* gene underwent trans-splicing, with its 3′-end duplicated in the IR region, while its 5′-end was present in the LSC region.

3.2. Comparative cp genome analyses of Hydrangeaceae

We conducted a comparative analysis of the cp genomes of 12 Hydrangeeae and 22 Hydrangeaceae species, in addition to the 28 genomes utilized in this study. The average genome size of the Hydrangeeae species was determined to be 157,592 bp, slightly larger than that of the Hydrangeaceae species, which had an average size of 157,413 bp (Fig. 2**A-i**). Similarly, the size of the LSC region in the Hydrangeeae species was measured at 86,594 bp, while in the Hydrangeaceae species, it was slightly larger, at 86,769 bp (Fig. 2**A-ii**). Additionally, both Hydrangeeae and Hydrangeaceae encoded the same size SSC region (18,680 bp) in their genomes (Fig. 2**A-iii**). In contrast, the IR region in the Hydrangeeae species (26,155 bp) was found to be slightly longer than that of the Hydrangeaceae species (26,119 bp) (Fig. 2**A-iv**). Both Hydrangeeae and Hydrangeaceae exhibited a similar GC content of 37.8% (Fig. 2**A-v**).

3.3. Analysis of Hydrangea cp genome boundary regions

The borders of the SC and IR regions in the cp genomes of *Hydrangea* were compared with other genomes, and according to the gene locations, we categorized the SC and IR region borders into three types (Supplementary Fig. S2). The boundaries of these regions exhibited a high level of consistency within the genus. In Type A, *rps19* was found to straddle the border between the LSC and IRb regions in most of the *Hydrangea* cp genomes (Supplementary Fig. S2A). However, there were exceptions in *H. quercifolia*, *H. caerulea*, *H. viburnoides* and *Deutzia paniculata*. In Type B, the *rps19* gene in *H. quercifolia*, *H. caerulea* and *D. paniculata* completely shifted to the LSC region (Supplementary Fig. S2B). In Type C, both the *rps19* and *rpl22* genes of *H. viburnoides* extended into the IR region, leading to the shift of the *rps3* gene from the LSC into the boundary region between the LSC and IRb (Supplementary Fig. S2C).

Additionally, in Types A and C, at the boundary regions between IRb and SSC, the pseudogene *ycf1* and the *ndhF* gene overlapped in all the cp

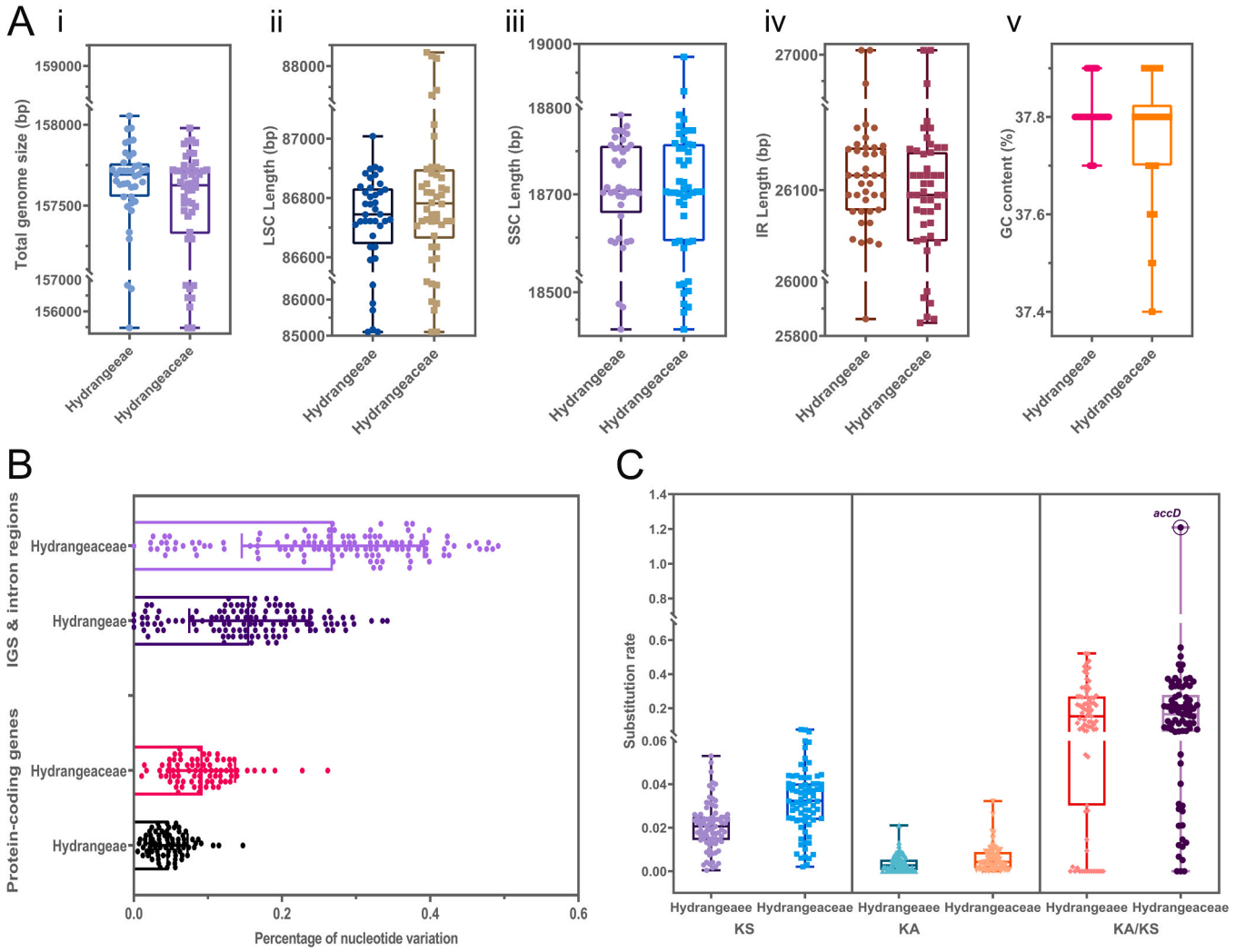

Fig. 2. Comparison of both Hydrangeeae and Hydrangeaceae cp genomes. **A**. (i) Total genome size. (ii) LSC region length. (iii) SSC region length. (iv) IR region length. (v) GC content. **B**. Nucleotide variation of protein-coding genes and intergenic regions and intron regions between Hydrangeeae and Hydrangeaceae. **C**. The K_S, K_A and K_A/K_S nucleotide substitution values of Hydrangeeae and Hydrangeaceae.

genomes of *Hydrangea*, except in *H. quercifolia*, in *H. caerulea* and *D. paniculata*. In these three genomes, the pseudogene *ycf1* shifted to the IR region, while the *ndhF* gene was relocated to the SSC region. Conversely, the genes situated in the SSC/IRa and IRa/LSC regions exhibited a notable degree of conservation throughout the cp genome of the *Hydrangea* genus.

3.4. Identification of divergent sequence hotspot regions in the Hydrangeaceae chloroplast genomes

To examine the extent of sequence polymorphism, we employed the mVISTA and DnaSP6 programs to assess the genetic distinctions within the complete cp genomes of species within the Hydrangeeae and Hydrangeaceae families. During mVISTA analysis, we observed that the variation in the IR regions of the Hydrangeeae cp genome was notably less than that in the single-copy regions (Supplementary Fig. S3). In the coding region, including *accD*, *atpF*, *matK*, *ndhI*, *psbC*, *rbcL*, *rpoC2*, *ycf1* and *ycf2,* substantial differences were identified, indicating distinct evolutionary patterns for these genes. Additionally, intergenic sequence (IGS) regions with pi values greater than 0.25, such as *trnK–rps15* (0.284), *psbK–psbI* (0.261), *psbI–trnS* (0.264), *trnG–trnfM* (0.336), *rps4–trnT* (0.259), *trnT–trnL* (0.342), *ndhC–trnV* (0.257), *petD–rpoA* (0.282), *rpl22–rps19* (0.266), *ndhF–rpl32* (0.286), *rpl32–trnL* (0.275), *ccsA–ndhD* (0.32), *ndhE–ndhG* (0.295), *ndhG–ndhI* (0.268) and *rps14–ycf1* (0.297), displayed high variability.

Concerning nucleotide variation (pi) in the protein-coding genes of Hydrangeeae and Hydrangeaceae, the ranges were 0.007–0.147 and 0.014–0.262, respectively, with average values of 0.047 and 0.093 (Fig. 2**B**). Similarly, the pi values for introns and intergenic regions of both Hydrangeeae and Hydrangeaceae ranged from 0 to 0.336 and 0–0.492, respectively, with average values of 0.156 and 0.269 (Fig. 2**B**). Additionally, within Hydrangeeae, the average pi value was 0.153 for the SSC region, 0.107 for the LSC region, and 0.024 for the IR region (Fig. 2**B**). Likewise, within Hydrangeaceae, the average pI value was 0.248 for the SSC region, 0.192 for the LSC region, and 0.048 for the IR region (Fig. 2**B**).

3.5. Positive selection analysis in the Hydrangeaceae cp genomes

The ratio of nonsynonymous (K_A) to synonymous substitution (K_S), K_A/K_S, was calculated for 79 protein-coding genes in both the Hydrangeeae and Hydrangeaceae cp genomes (Fig. 2**C**). The results indicated that all protein-coding genes of Hydrangeeae displayed very low K_S values (<0.05), with an average of 0.021. Similarly, the K_A values were also comparatively low (<0.02), with an average of 0.004. The K_A/K_S substitution ratio of all protein-coding genes ranged from 0 to 0.522, with an average of 0.173. The substitution analysis was expanded across all Hydrangeaceae genomes to gain a comprehensive understanding of the substitution patterns in these genomes. The K_S and K_A substitution rates ranged from 0.002 to 0.075 and 0.0002–0.032, respectively, with average values of 0.033 and 0.006 (Fig. 2**C**). The K_A/K_S substitution ratio revealed that except for *accD* (with a K_A/K_S ratio of 1.209), all of the other 78 protein-coding genes exhibited relatively low substitution values (< 0.6), with an average value of 0.192. The K_A/K_S ratio of *accD* in Hydrangeeae and Hydrangeaceae ranged from 0 to 1.818 and 0–5.0286, respectively.

3.6. Analysis of accD

The *accD* genes of all Hydrangeaceae species were compared and identified to be identical copies of 18 bp tandem repeats in the Hydrangea II clade excluding *H. arguta*. We categorized the species into three groups depending on whether they had a tandem repeat in the acetyl-CoA carboxylase beta subunit D domain of *accD* (Fig. 3**A**). *H. arguta* (Group A) did not exhibit any copy of the 18 bp tandem repeat in *Hydrangea* clade II. Group B comprised five *Hydrangea* species (*H. scandens*, *H. luteovenosa*, *H. serrata*, *H. macrophylla* and *H. aspera*) and one *H. febrifuga* sample that contained two copies of the 18 bp tandem repeat. Group C included *H. febrifuga*, *H. daimingshanensis* and *H. hirsuta* with four copies of the 18 bp tandem repeat (Fig. 3**B**). The protein reading frame was not disrupted in any of the species due to the repeated region being comprised of six amino acid codons (Fig. 3**C**). This event occurred in the common ancestor of the sections *Macrophyllae*, *Chinenses* and *Dichroa* (Fig. 3**D**).

We analyzed the protein structure of accD in each of the three groups of genomes. For each group, one representative species was selected. The accD protein structure was successfully retrieved from the AlphaFold database using the id A0A481XL48.1, which showed a sequence identity of 92.7% with the structure A_9ERIC. To ensure the reliability of the obtained protein structures, we conducted several validation tests demonstrating the quality of the structures, with the Ramachandran plot indicating that 81.7% of residues were in the most favored region, 10.6% in the additional allowed region, 4.5% in the generously allowed region, and 2.0% in the disallowed region. Moreover, the clash score revealed that only 0.25% of atomic coordinates had unfavorable interactions, confirming a lack of significant clashes in the protein structure. Additionally, the all-atom structural annotation obtained from MolProbity showed a root mean square deviation of 1.57 Å. The overall quality factor of the protein structures was assessed using the ERRAT, ProSA, and ProQ scores, which were found to be 95.3%, − 7.11 and − 0.3, respectively. These scores fell within statistically significant ranges, indicating high-quality protein structures [59]. We considered specific insertions in regions 241–246 and 241–259 of the annotated structures for further analysis. These insertions were investigated to decipher potential conformational changes related to functional activity.

The presence of tandem repeats in the accD protein, specifically at positions 241–246, resulted in the introduction of extended α-helices in the C-terminal region (516−526) and loops in the central region (260−270) (Fig. 3**D, E**). On the other hand, the insertion at position 241–259 led to the loss of α-helices in the C-terminal region (511−515) and the appearance of extended coils (260−266) in the central region of the protein (Fig. 3**D, E**).

According to the protein structure analysis, we conducted qPCR experiments to analyze the expression pattern of *accD* in the three groups. One representative species from each group was chosen, with Group A serving as the control. The qPCR analysis revealed that the CT value of *H. paniculata* (Group A) was 27.80933 ± 1.203. In contrast, the CT values of *H. febrifuga* (Group B) and *H. hirsuta* (Group C) were 19.181 ± 0.157 and 22.80767 ± 0.084, respectively (Supplementary Fig. S4).

3.7. Analyses of amino acid abundance and codon usage in the Hydrangeaceae cp genomes

In the cp protein-coding genes of the 50 species within Hydrangeaceae, a total of 62 codons were identified, with the protein-coding genes encompassing a range of 25,777 (*H. densifolia*) to 27,012 (*H. quercifolia*) codons. We carried out a comparative analysis of amino acid distribution within the protein-coding genes of both the Hydrangeeae and Hydrangeaceae cp genomes (Supplementary Fig. S4). Within these codons, leucine (L) was the most abundant amino acid at 10.475%, followed by isoleucine (I) at 8.616%, while cysteine (C) was the least prevalent amino acid, accounting for only 1.159% (Supplementary Fig. S5). Notably, all amino acids were represented by a minimum of two synonymous codons, except for tryptophan (UGG), which had the lowest codon usage preference at 1.842% since it had only one coding codon. To further investigate synonymous codon preference, we calculated the relative synonymous codon usage (RSCU) values, which were categorized into three groups: high preference (RSCU > 1.0 – denoted with green); no preference (RSCU = 1.0 – denoted with black) and low preference (RSCU < 1.0 – denoted with red) (Supplementary Fig. S6). The analysis identified 16 codons with RSCU values greater than 1.0. Of

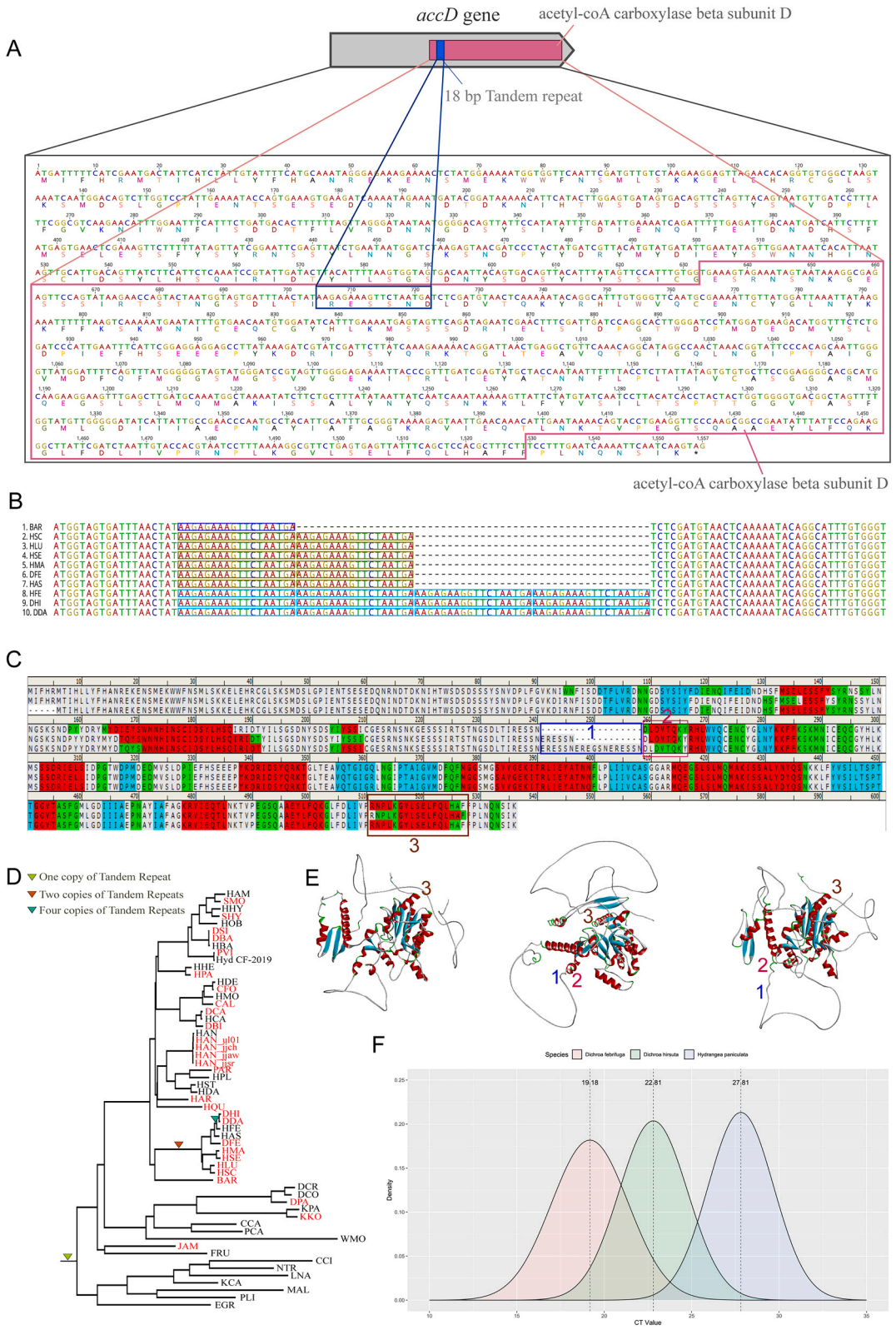

Fig. 3. Analysis of *accD* gene. **A**. The gray rectangular arrow box depicts the overview of *accD* gene; the box within the *accD* gene indicates the acetyl-CoA carboxylase beta subunit conserved domain (coral pink), and the blue box indicates the location of 18 bp tandem repeat. **B**. Comparison of *accD* gene in *Hydrangea* clade II; the rectangular box indicates the distribution of 18 bp tandem repeats in *Hydrangea* clade II. **C**. Amino acid sequences of *accD* gene from three representative groups. **D**. The phylogenetic tree shows the presence of single, two and four copies of 18 bp tandem repeats. **E**. Sequential and structural differentiation of the accD protein. The rectangular box 1 represents the tandem repeats; the rectangular boxes 2 and 3 denote the structural aberration of secondary structural elements after insertion in the region. Red, helices; turquoise, sheets; gray, loops; green, turns. **F**. Quantitative PCR (qPCR) analysis of *accD* gene expression in the species *Hydrangea petiolaris* (control, one copy of tandem repeat), *H. febrifuga* (two copies of tandem repeats) and *H. hirsuta* (four copies of tandem repeats).

these, 14 were codons ending in A or T. Among the A/T-ending codons, 37.5% terminated with A, while the remaining 62.5% ended with T. Conversely, there were 43 codons with RSCU values less than 1.0, and 30 of them concluded with G or C. Among the G/C-ending codons, 50% were C-ending, and the other 50% were G-ending.

3.8. RNA editing

We employed PREPACT to predict RNA editing modifications in 79 protein-coding genes of 57 Hydrangeaceae and closely related species. Out of the 79 protein-coding genes analyzed, 22 protein-coding genes did not exhibit any RNA editing process. However, in the remaining 57 protein-coding genes, all the species displayed a consistent pattern of C to U RNA editing sites within their sequences (Supplementary Fig. S7). The majority of RNA editing events were observed in the gene *ycf1*, with 39–49 editing sites accounting for 15.02–19.522% of the total edits. Additionally, significant numbers of editing sites were identified in *ycf2* (26–30 editing sites, 9.40–12.39%), *matK* (10–28 editing sites, 3.968–11.067%) and *accD* (8–16 editing sites, 2.35–5.46%) (Supplementary Fig. S7).

3.9. Repeat sequence analyses

We utilized REPuter to analyze the presence of interspersed repetitive sequences with a repeat unit length exceeding 30 bp in the cp genomes of 50 Hydrangeaceae species. We identified four types of repeats, including forward repeats (F), palindromic repeats (P), inverted repeats (R) and complementary repeats (C) (Fig. 4). The results revealed variations in the number of repeats, ranging from 18 (*H. petiolaris*) to 30 (*H. crassifolia*). Specifically, the repeat analysis showed 11–20 forward duplications, 5–9 palindromic repeats, and 0–1 inverted duplications in the Hydrangeaceae species. Notably, *H. moellendorffii* and *H. crassifolia* each contained one complementary repeat, which was absent in the remaining species. Most of the repeats observed were forward repeats, with sizes ranging from 30 to 39 base pairs.

Additionally, we conducted an analysis of simple sequence repeats (SSRs) within the cp genomes of the 50 Hydrangeaceae species. The total number of SSRs ranged from 40 in *H. hydrangeoides* to 71 in both *H. petiolaris*_JJAR and *H. petiolaris*_JJSR. Among these SSRs, A/T mononucleotide repeats were the most prevalent, constituting 76.608% of the total (Supplementary Fig. S8). We identified six types of SSRs, encompassing mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats. Mononucleotide SSRs exhibited variations in the number of repeat units, ranging from 31 to 59, while dinucleotide SSRs varied from two to eight units, and trinucleotide SSRs ranged from one to five units (Supplementary Fig. S9). Notably, *H. petiolaris* displayed a higher number of SSRs than other *Hydrangea* species. The first four types of SSRs collectively represented 98.78% of the identified repeats. Furthermore, we observed that a significant majority of the SSRs were situated in the LSC region of the cp genome (83.02%), followed by the SSC region (13.64%), and finally the IR regions (3.34%) (Supplementary Fig. S10).

3.10. Phylogenetic analysis

In the phylogenomic studies, we utilized at least one species each from the *Schizophragma*, *Decumaria*, *Pileostegia*, *Heteromallae*, *Cardiandra*, *Deinanthe*, *Calyptranthe*, *Asperae*, *Hydrangea*, *Dichroa*, *Chinenses*, *Macrophyllae*, and *Broussaisia* sections of the Hydrangeeae tribe. Consequently, we sequenced 22 Hydrangeeae species (25 individuals), two *Philadelphia* species and one Jamesioideae species as the outgroup species, in addition to 29 species retrieved from NCBI. For the present study, we employed three datasets for ML tree construction, including (i) whole cp genome sequences (WCP; Supplementary Fig. S11), (ii) 79 concatenated protein-coding genes (PCS; Supplementary Fig. S12) and (iii) translated amino acid sequences (AA; Supplementary Fig. S13), along with one Bayesian tree using PCS (Bayesian PCS; Fig. 5), to infer the phylogenetic position of the *Hydrangea* genus within the Hydrangeaceae family.

We obtained similar topologies across all the datasets, WCP, PCS, AA, and Bayesian PCS, with slight variations in bootstrap values for all phylogenomic analyses. All the phylogenomic trees displayed paraphyly within the Hydrangeeae tribe (Fig. 5; Supplementary Figs. S11-S14).

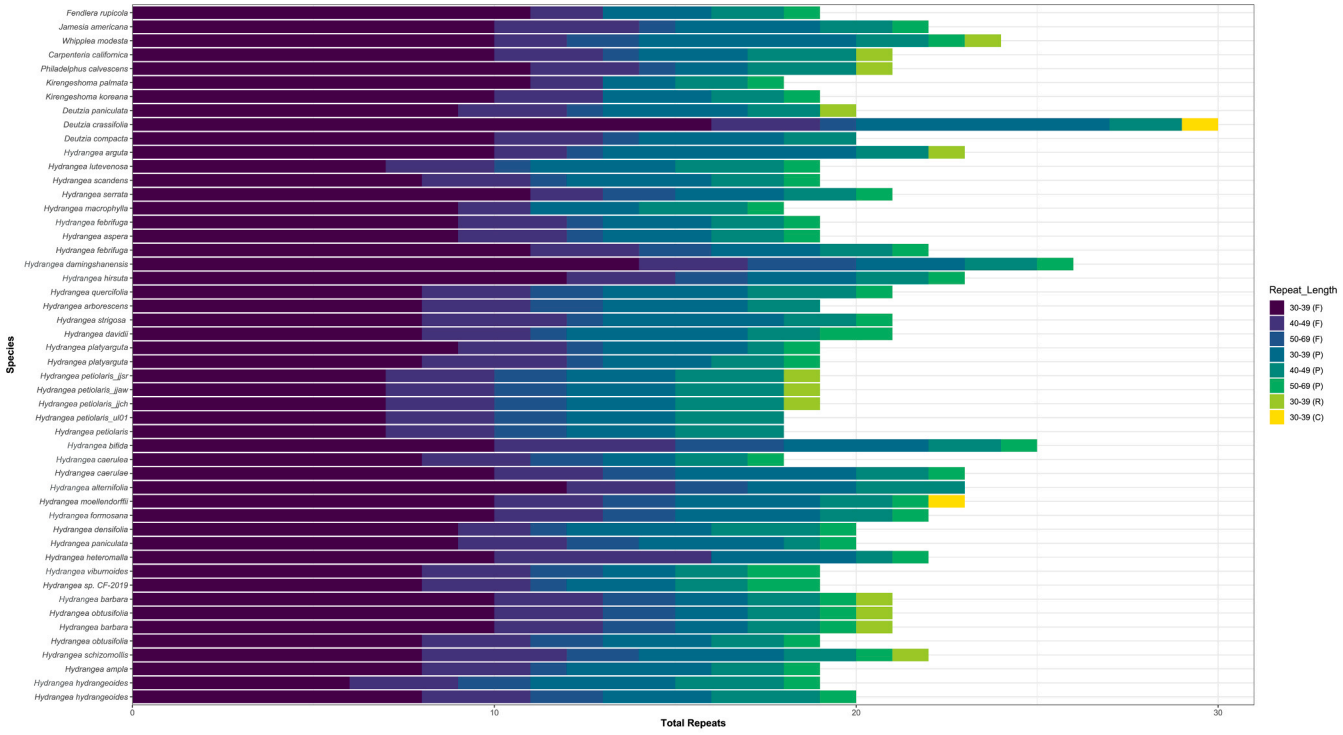

Fig. 4. The distribution of complementary (C), forward (F), palindromic (P) and reverse (R) repeats in the Hydrangeaceae cp genomes.

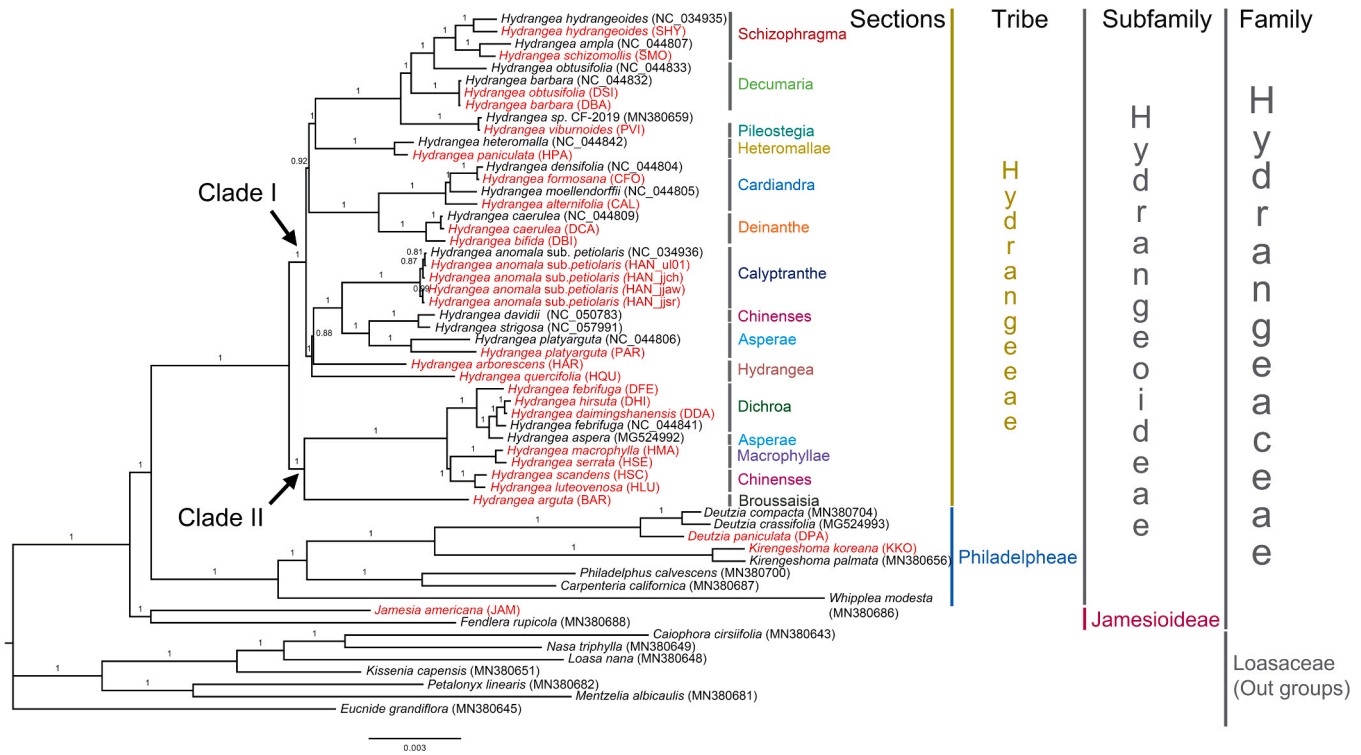

Fig. 5. Bayesian Markov chain Monte Carlo (MCMC) inference of the MrBayes phylogenetic tree. The gamma model of rate variation and the HKY85 substitution model were utilized for this analysis. The Bayesian tree is depicted with the Bayesian inference posterior probability value given for each node.

Within the Hydrangeeae tribe clade, all sections formed a monophyletic group, except for three sections, namely, *Asperae*, *Chinenses* and *Decumaria,* which remained nonmonophyletic in all phylogenomic analyses. The section *Broussaisia* was identified as the basal group in the Hydrangeeae tribe. Most sections showed higher bootstrap values, except for *Calyptranthe*, *Chinenses*, and *Macrophyllae*.

Across all datasets, the 33 *Hydrangea* species could be divided into two clades: *Hydrangea* I, consisting of 23 species (30 individuals), and *Hydrangea* II, consisting of 10 species. The *Hydrangea* I clade comprised the sections *Schizophragma*, *Decumaria*, *Pileostegia*, *Heteromallae*, *Cardiandra*, *Deinanthe*, *Calyptranthe*, *Asperae* and *Hydrangea,* while the *Hydrangea* II clade consisted of the sections *Dichroa*, *Chinenses*, *Macrophyllae* and *Broussaisia*. Within Clade I, the sect. *Decumaria* contained two species (*H. obtusifolia* and *H. barbara*) that exhibited monophyly with strong bootstrap values. Notably, within the *Hydrangea* I clade, the species *H. davidii* (NC_050783) belonging to the *Chinenses* section clustered with *H. strigosa* (Asperae section) with a strong bootstrap value. On the other hand, within the *Hydrangea* II clade, the species *H. aspera* (MG524992) from the *Asperae* section clustered with *H. febrifuga*.

Furthermore, we collected four samples of *H. petiolaris*, with three samples from Jeju Island and one sample from Ulleung Island (Korea), and one sample retrieved from NCBI (Ulleung Island). According to the phylogenomic analyses, these five individual samples were divided into two groups: two samples (UL) from Ulleung Island + one sample (JJCH) from Jeju Island (Mt. Halla) and two samples (JJAW and JJSR; excluding Mt. Halla) from Jeju Island. However, the bootstrap values of UL01 and JJCH were observed to be very low in all phylogenetic analyses.

3.11. Divergence time analysis

Utilizing sequence data from 79 cp protein-coding genes, our divergence time analysis indicated that the initial divergence of the Hydrangeaceae family stem lineage occurred approximately 59.49 million years ago during the early Paleocene epoch. This estimation had a 95% highest probability density (HPD) range of 48.33–71.9 million years ago (Fig. 6**A**). Within the subfamily of Hydrangeoideae, the Philadelpheae and Hydrangeeae tribes diverged at approximately 56.49 Ma, with a 95% HPD of 45.7–68.21 Ma.

The disjunction between the Ulleung Island (UL01) and Jeju Island (Mt. Halla; JJCH) samples of *H. petiolaris*, as well as between the Jeju Island (JJAW and JJSR) samples of *H. petiolaris*, occurred in the Pleistocene epoch and was estimated to be 0.27 Ma (95% HPD = 0.08–0.65 Ma). The crown nodes of UL01 and JJCH were estimated to have formed approximately 0.13 Ma (95% HPD = 0.03–0.31 Ma). Similarly, the crown nodes of JJSR and JJAW were likely to have originated approximately 30,000 years ago (95% HPD = 0–0.16 Ma).

3.12. Phylogenetic informativeness

The per-site phylogenetic informativeness (PI) profiles were determined for 79 protein-coding genes obtained from 57 species, including those from the Hydrangeaceae family and closely related species, using PhyDesign software. Among all the protein-coding genes, the *ycf1* gene displayed the maximum per-site PI, followed by *ndhF*, *rps15*, *matK*, *rpoC2*, *rpoB*, *ccsA*, *psbK*, *ycf2*, *psbM*, *ndhK*, *rpl22*, *rpl20*, *ndhD*, *rpl33*, *rpoA*, *rps16* and *accD* (Fig. 6**B**).

3.13. Ancestral area reconstruction

The biogeographical history of the Hydrangeaceae family was inferred using various ancestral area reconstruction methods, including S-DIVA (Supplementary Fig. S14), DEC, S-DEC, BBM, BioGeoBEARS (BGB), and BayArea (BA) analyses. The results of the DEC (Fig. 7), S-DEC (Supplementary Fig. S15), and BGB (Supplementary Fig. S16) analyses suggested that the origin of Hydrangeaceae was in America, with North America (C) and South America (D) being the potential areas of origin, followed by numerous events of vicariance and dispersal into the Asia (A) region. However, the BBM (Supplementary Fig. S17) and S-DIVA

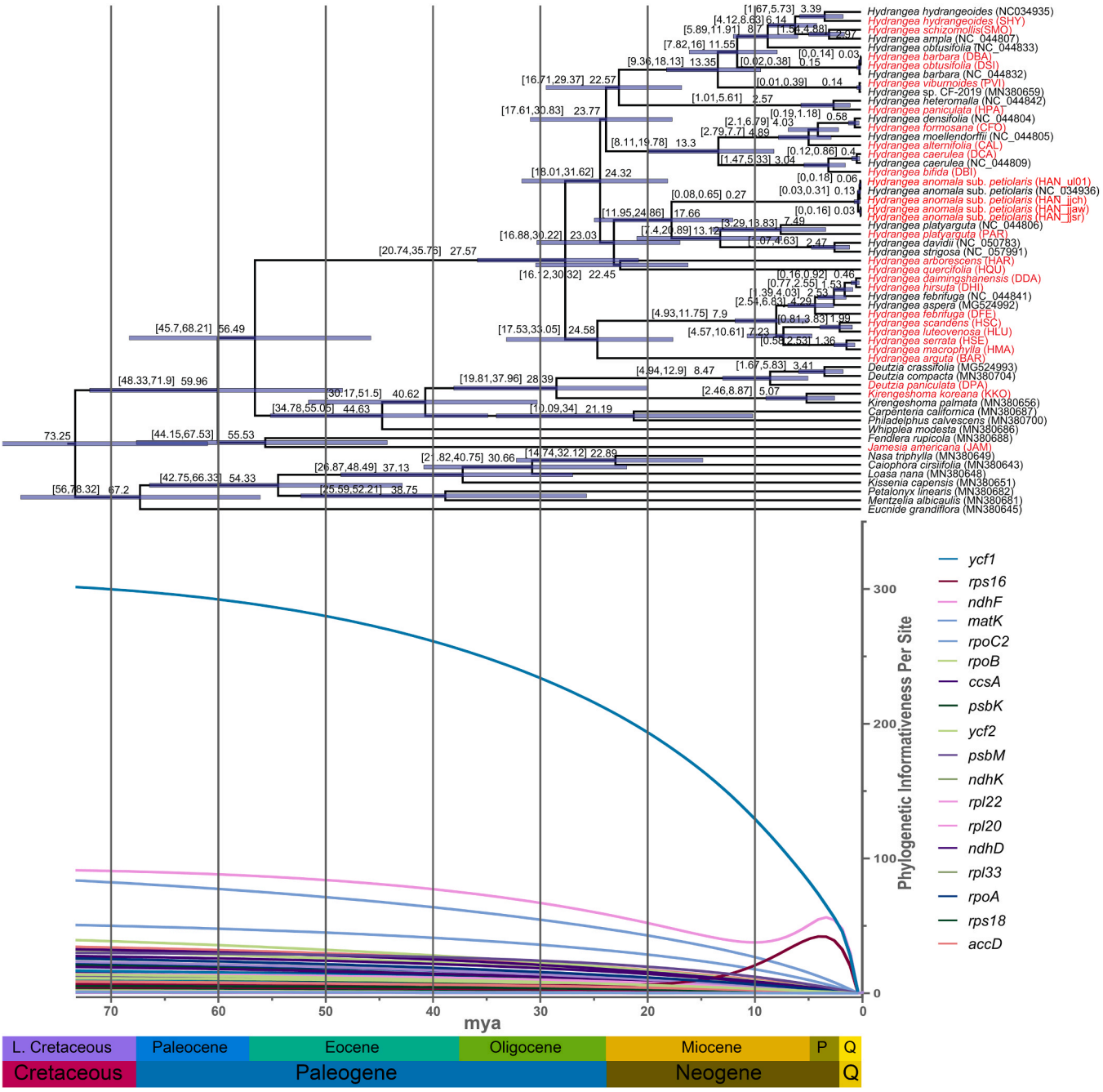

Fig. 6. **A.** The BEAST program is used to construct a molecular clock tree. The distribution of age estimates (in millions of years) is shown in the branches. Blue bars represent the 95% posterior density creditability intervals for node ages. **B.** Net phylogenetic informativeness plotted against the ultrametric phyletic tree constructed on 79 protein-coding genes.

(Supplementary Fig. S14) analyses presented different scenarios. BBM suggested the possibility of the origin being in North America (C) and Asia & North America (AC), while S-DIVA indicated that Asia was the likely place of origin for the Hydrangeaceae family (Supplementary Fig. S14). Regarding the Hydrangeoideae subfamily, both S-DIVA and BBM analyses consistently pointed to Asia as the origin of the *Hydrangea* genus (Supplementary Figs. S14 and S17). In contrast, DEC, S-DEC, BGB and BayArea analyses yielded somewhat conflicting results, suggesting that *Hydrangea* could have originated in Asia or North America (Fig. 7, Supplementary Figs. S15-S16 and S18).

Further in-depth analyses revealed variable origins for the genus *Hydrangea*. The S-DIVA (Supplementary Fig. S19), DEC (Supplementary Fig. S20) and BBM (Supplementary Fig. S22) analyses pointed to the possibility of East Asia (A) as a potential area of origin. On the other hand, BGB analysis indicated the possibility of a combined ancestral area comprising East Asia (A) and Hawaii (M) (Supplementary Fig. S21). Interestingly, the S-DEC (Fig. 8) and BA (Supplementary Fig. S23) analyses proposed an ancestral area that includes East Asia (A) and Southeast North America (F) as the possible origin for the genus *Hydrangea*.

4. Discussion

The chloroplast (cp) genome of angiosperms has played a crucial role in studying phylogeny and analyzing evolutionary relationships [41]. The comparative analysis of plastomes has found widespread

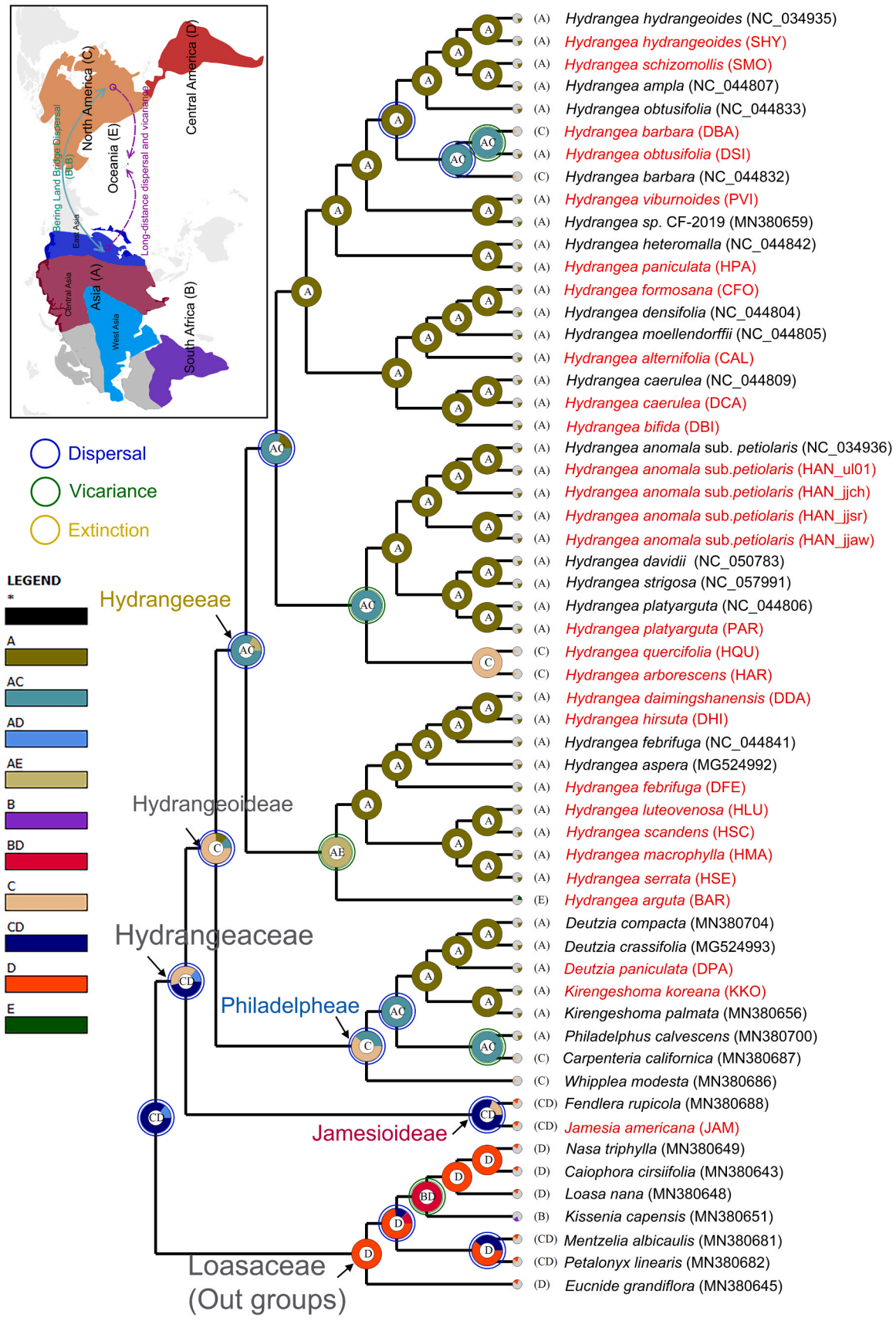

Fig. 7. The DEC method for inferring ancestral geographical ranges within the Hydrangeaceae family relies on utilizing simplified BEAST and MrBayes phylogenetic trees. At various nodes in the phylogeny, pie charts are employed to visually convey the relative likelihoods of different ancestral area reconstructions. Additionally, a rectangular diagram is used to depict the potential dispersal pathways of Hydrangeeae species during the late Oligocene epoch, with geographical regions represented by letters: A, Asia; B, South Africa; C, North America; D, Central America; and E, Oceania.

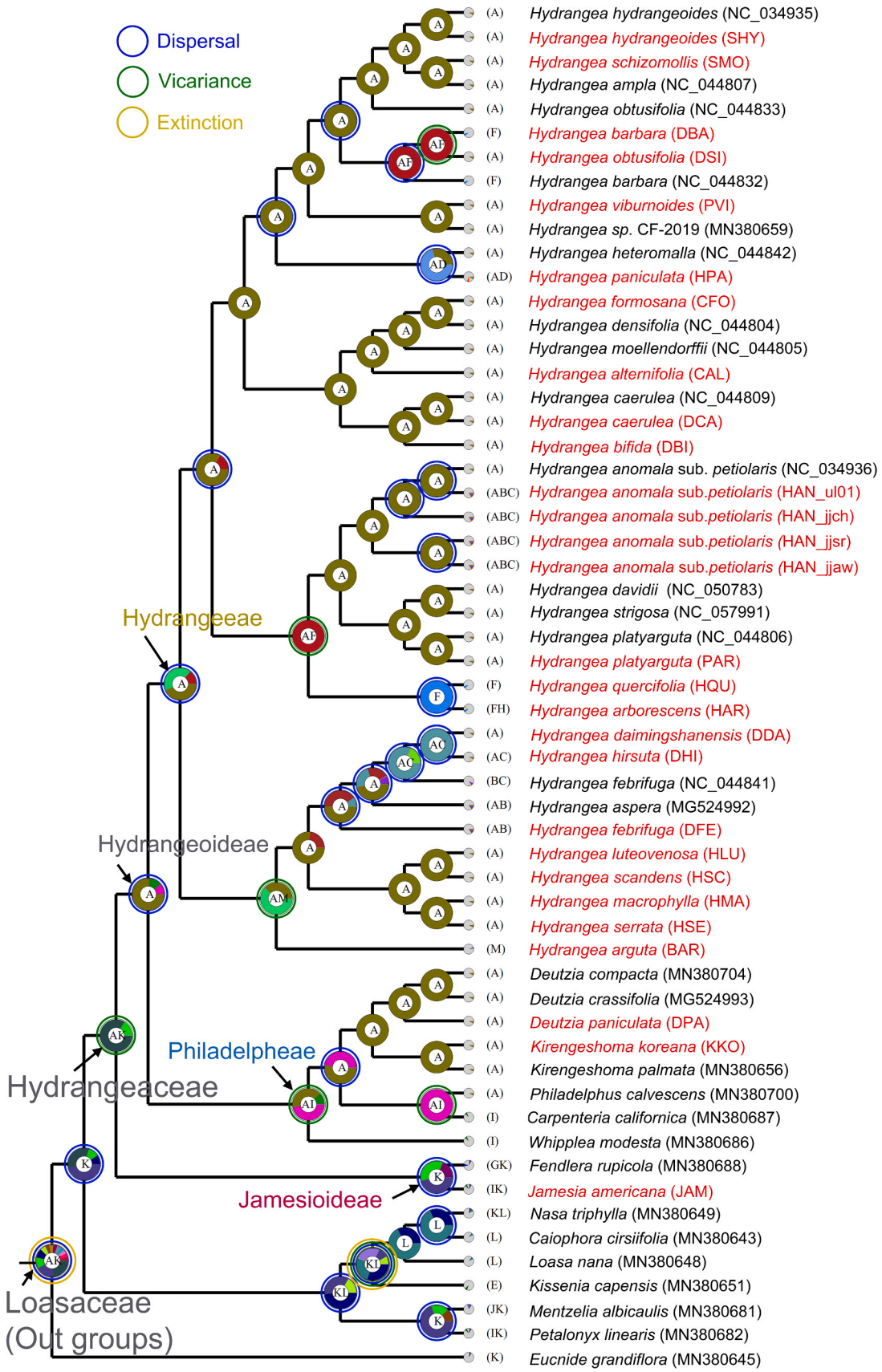

Fig. 8. The in-depth S-DEC method for inferring ancestral geographical ranges within the Hydrangeaceae family relies on utilizing simplified BEAST and MrBayes phylogenetic trees. At various nodes in the phylogeny, pie charts are employed to visually convey the relative likelihoods of different ancestral area reconstructions. A, eastern Asia; B, southern Asia; C, southeastern Asia; D, the Asian part of Russia; E, South Africa; F, southeastern North America; G, southwestern North America; H, northeastern North America; I, western North America; J, Canada; K, Mexico and Guatemala; L, Peru and Chile; M, Hawaii.

application across various taxa, including the Hydrangeaceae family within the Cornales order [18]. Previous studies have compared 13 cp genomes of *Hydrangea* species to establish phylogenetic relationships within the Cornales order [18]. However, this study signifies the first comprehensive comparative analysis within the Hydrangeaceae family, focusing on the Hydrangeeae tribe. In the present study, we sequenced 25 individuals from 22 *Hydrangea* species and used three outgroup species, two from Philadelpheae and one from Jamesioideae, to understand the phylogenomic relationship within the *Hydrangea* genus (Fig. 1).

In summary, our study unveiled consistent patterns in gene arrangements and the overall gene count within the plastomes of both the *Hydrangea* genus and the broader Hydrangeaceae family within the Cornales order (Table 1) [18,27,43]. We emphasize the conservation of coding regions compared to noncoding regions and the relatively higher conservation of IR regions compared to single-copy regions in both Hydrangeeae and Hydrangeaceae cp genomes (Fig. 2**B**). Notably, the intergenic sequence (IGS) regions, such as *trnK–rps15*, *psbK–psbI*, *psbI–trnS*, *trnG–trnfM*, *rps4–trnT*, *trnT–trnL*, *ndhC–trnV*, *petD–rpoA*, *rpl22–rps19*, *ndhF–rpl32*, *rpl32–trnL*, *ccsA–ndhD*, *ndhE–ndhG*, *ndhG–ndhI* and *rps14–ycf1,* exhibited high variability (Fig. 2**B**, Supplementary Fig. S3). These discoveries offer valuable insights into the genetic diversity and evolutionary characteristics of cp genomes within Hydrangeeae and Hydrangeaceae species. Understanding these patterns may enable the application of highly variable regions in intergenic sequences for DNA barcode encoding and phylogenetic analysis of Hydrangeeae [53].

The substitution analysis revealed relatively low substitution values in 78 protein-coding genes. Our study demonstrated that K_S was significantly higher than K_A, indicating a relatively low evolutionary rate for both Hydrangeeae and Hydrangeaceae species compared to other angiosperm lineages. Moreover, the K_A/K_S ratio was found to be very low for most genes, suggesting that purifying selection is the dominant force maintaining the functionality and sequence conservation of these genes (Fig. 2**C**). However, *accD* displayed a higher K_A/K_S ratio, indicating a different evolutionary pattern and possible functional implications. In the common ancestor of the *Hydrangea* clade II sections *Macrophyllae*, *Chinenses* and *Dichroa*, excluding the species *H. arguta*, an interesting event occurred involving the insertion of an identical 18 bp tandem repeat sequence (Fig. 3**B, D**). Notably, the *Macrophyllae* and *Chinenses* sections had two copies of the 18 bp tandem repeat, while the *Dichroa* section had four copies, except *H. febrifuga* (two copies). This synapomorphic character might be attributed to replication slippage during DNA replication. Despite the insertion of 36 and 72 bp tandem repeats in the *accD* gene, the reading frame was not disrupted in any of the species (Fig. 3**C**), and the GC content of these repeats was very low (27.8%). Tandem repeat sequences were also identified in the *accD* and *rpl2* genes of other angiosperms (*Pisum sativum*, *Lathyrus sativus*, *Capsicum annuum* and *Medicago truncatula*) and gymnosperms (cupressophytes) [25,31,47,50]. However, the repetitive elements in these species exhibited no similarity, indicating a species-specific nature. However, the occurrence of tandem repeats in angiosperms did not occur in the common ancestor [47]. In contrast, in the present study, we identified that the synapomorphy feature occurred in the common ancestor of *Hydrangea* clade II except *H. arguta*, similar to cupressophytes [47]. Structural analyses validated the presence of tandem repeats in *accD* and revealed disruptions in folding conformations, potentially affecting the protein's functional activity by altering α-helices in both the N- and C-termini due to the replication slippage mechanism (Fig. 3**C, E**).

Furthermore, qPCR analysis confirmed slightly lower expression levels in species with two (CT value = 19.181 ± 0.157) and four copies (22.80767 ± 0.084) of tandem repeats compared to the species with a single-copy tandem repeat (27.80933 ± 1.203) (Supplementary Fig. S4). This decrease in gene expression might be linked to the loss of α-helices in the C-terminal region (511−515). Despite these structural changes, the acetyl-CoA carboxylase beta subunit D domain of the *accD* gene in *Hydrangea* clade II is presumed to be functional, and the repetitive elements observed in these species could potentially have a regulatory role in modulating protein function.

The exploration of codon usage bias has aided in the comprehension of the interplay between the cp and nuclear genomes [85]. Previous studies have shown that codons encoding leucine and isoleucine occur most frequently in the cp, whereas codons for cysteine are the least frequent [2,60,68,85]. This feature is consistent in the cp genomes of all Hydrangeaceae and the closely related Loasaceae species (Supplementary Figs. S4 and S5). These results are consistent with other angiosperm cp genomes, with most codons showing a higher A/T preference in the third position [60,65,71,74]. Our analysis also demonstrated that codons ending with A or T at the 3′ end have a higher frequency of occurrence than those ending with C or G (Supplementary Fig. S6). Codons ending with A/T at the 3′ position typically exhibit an RSCU value greater than 1, while codons ending with C or G at the 3′ position have an RSCU value ≤ 1. This pattern has been observed in other angiosperms [60,71], and it may be linked to the prevalence of A or T in the IR region [7,60].

RNA editing is a posttranscriptional modification process that encompasses an indel or the conversion of cytidine (C) to uridine (U) nucleic acid bases within the plastids of angiosperms [19]. In this study, it was observed that the *ycf1* gene exhibited the highest number of editing sites among the plastid genomes of the Hydrangeaceae family (Supplementary Fig. S7). This observation may be attributed to the length of the gene. The identified RNA editing sites exclusively involved C-to-U conversions and affected a single site per editing event. Notably, all modifications appeared in the first or second nucleotides of the codon [60]. Among the diverse amino acid conversions attributed to RNA editing sites, the change from serine to leucine represented one-third of the total conversions [71]. Notably, all species shared similar RNA editing sites, indicating that these sites likely originated in a common ancestor before the species diverged. In general, the codon preferences of RNA editing targets and the inclination toward greater protein hydrophobicity and site distribution exhibited parallel trends throughout the Hydrangeaceae family.

The repeat units were found to be widely distributed and significantly contributed to the evolution of cp genomes [71]. Among the fifty Hydrangeaceae plastomes examined, the number and types of repeat units varied (Fig. 4). Prior research has reported that the variability in repeat number and diversity plays a pivotal role in shaping the structural organization of plastomes. However, these large repeat regions did not exhibit any correlation with rearrangement endpoints [11]. Furthermore, the distribution of SSRs within the plastome is considered a valuable marker for population genetics and phylogenetic analysis [73]. In our study, the number of SSRs detected across the 50 Hydrangeaceae species ranged from 40 to 71, with the LSC region displaying a higher density of SSRs in comparison to the IR and SSC regions (Supplementary Fig. S10). This observation is consistent with earlier research on cp genomes in angiosperms [71]. The SSR analysis conducted in this study revealed that single-nucleotide SSRs, particularly A/T repeats, had the highest abundance among the 50 Hydrangeaceae species (Supplementary Fig. S8), with a total count of 2073. Mono-, di-, tri-, and tetranucleotide repeats collectively accounted for 98.78% of all SSRs, while penta- and hexanucleotide repeats were rare (Supplementary Fig. S9). Notably, the plastomes of these 50 Hydrangeaceae species exhibited a higher AT content than GC content, and SSRs exhibited a strong AT bias, a common characteristic observed in the plastomes of higher plants [38,44]. Repetitive sequences play a crucial role in generating insertion and substitution mutations. Previous studies have provided evidence of widespread substitutions and deletions occurring in the LSC and SSC regions of plastomes [1,51].

Plastid genome sequences have proven effective in elucidating the phylogenetic relationships of plants across various taxonomic levels and within diverse taxa [30,46,5,52,55]. In prior research, the relationships

within the Hydrangeaceae family were elucidated using morphological traits and a limited set of molecular markers, leading to the classification of three major clades: Jamesioideae, Philadelpheae, and Hydrangeeae, as determined through phylogenomic analyses [13,66]. Earlier investigations have revealed that the Hydrangeeae tribe clade encompasses eight smaller genera that are nested within the polyphyletic larger genus *Hydrangea* [13,66]. [66] used five plastid markers but did not completely resolve the phylogenetic position of seven *Hydrangea* clades despite each clade being well supported by morphological data. [66] also identified two strongly supported clades within the tribe Hydrangeeae. The first of these, referred to as *Hydrangea* I, included *Cardiandra*, *Decumaria*, *Deinanthe*, *Pileostegia*, *Schizophragma*, and several members of the *Hydrangea* genus, albeit with some unresolved relationships. The second significant clade, termed *Hydrangea* II, comprised *Broussaisia* and *Dichroa*. [66] proposed that these clades might possess synapomorphic characters or unique combinations of morphological traits, some of which could be ancestral. [23] conducted a study testing the effectiveness of 13 plastid markers with a reduced sample size, aimed at resolving the core relationships within the tribe Hydrangeeae (excluding *Broussaisia*). Their findings suggested that the *Hydrangea* genus is not monophyletic, consistent with the conclusions of [66]. In 2015, [13] utilized four plastid and ITS sequences in an extensive dataset to ascertain the phylogenetic placement of the Hydrangeeae tribe. The analysis revealed that *H. arguta* clustered with the Hydrangea I clade, in contrast to the findings of [66] and [23]. Additionally, within *Hydrangea* I, *H. arborescens* and *H. quercifolia* were grouped in a weakly supported clade and were sisters to the rest of *Hydrangea* I. More recently, [18] used entire plastid genomes of 74 species from all families of the order Cornales to understand their phylogenetic relationships. Specifically, Fu et al. [18] used 13 species from nine genera of the tribe Hydrangeeae, and the results were consistent with those of [13]. In 2015, De Smet et al. introduced a new infrageneric classification proposal, cautioning that adopting a splitting approach to create numerous new genera within Hydrangeeae could introduce complexities into the taxonomy of the group. This could lead to either a proliferation of genera with only one species each or the emergence of multiple taxa with significant morphological variation, potentially making them difficult to identify reliably. As a result, they recommended a broader definition of *Hydrangea* that encompasses *Broussaisia*, *Cardiandra*, *Decumaria*, *Deinanthe*, *Dichroa*, *Pileostegia*, *Platycrater*, and *Schizophragma*. Therefore, in the present study, we configured all the genera of Hydrangeeae into the genus *Hydrangea*. Hence, we used 25 individuals from 22 species of the Hydrangeeae tribe in five different ML datasets and one Bayesian analysis to resolve their phylogenetic relationships. All the datasets showed consistent and similar topologies, and the tribe Hydrangeeae was classified into two major clades: *Hydrangea* I and *Hydrangea* II (Fig. 5; Supplementary S11-S13). The *Hydrangea* I clade comprised sections *Broussaisia*, *Macrophyllae*, *Chinenses* and *Dichroa*, while the *Hydrangea* II clade contained sections *Hydrangea*, *Asperae*, *Calyptranthe*, *Deinanthe*, *Cardiandra*, *Heteromallae*, *Pileostegia*, *Decumaria* and *Schizophragma*. In our analyses, the genera exhibiting morphological diversity were generally found to be monophyletic, except for sect. *Chinenses* and sect. *Asperae,* which were situated within the larger, polyphyletic *Hydrangea*. These outcomes were largely consistent with prior research, and the majority of the clades received robust support through strong bootstrap values [13,18,66]. In contrast, within clade I, sect. *Decumaria* comprised just two species and emerged as a monophyletic group with robust support in our phylogenomic investigation. This finding deviated from the results reported by [18] and [84]. Recently, [84] utilized 79 PCSs, and a phylogenetic study showed that sect. *Decumaria* formed a paraphyly. In addition, the sect. *Hydrangea*, *H. quercifolia* and *H. arborescens* formed the polyphyletic group and were only partly resolved (with low bootstrap values) by [18], corroborating earlier studies [13]. Additionally, *H. obtusifolia,* which belongs to sect. *Decumaria*, clustered with sect. *Schizophragma* with a strong bootstrap value (100%).

We used plastid genome datasets from Fu et al. [18] in our phylogenomic analyses. In our study, the species *H. obtusifolia* (NC_044833) was found to belong to sect. *Decumaria*, which clustered with sect. *Schizophragma,* and *H. aspera* (MG524992) was found to belong to sect. *Asperae* (Hydrangea I clade) and clustered with sect. *Dichroa* (Hydrangea clade II) (Fig. 5; Supplementary S11-S13). Similarly, *H. davidii* (NC_050783) belonged to sect. *Chinenses* (Hydrangea II) and clustered with sect. *Asperae* (Hydrangea I). These findings were consistent across all our analyses. To confirm these findings, we collected additional specimens of these three species and generated separate phylogenetic trees using the plastid genes *matK* and *rbcL* (Supplementary Figs. S24 and S25). Interestingly, we obtained better resolution and bootstrap values with the *matK* gene compared to *rbcL*. This is consistent with prior research, which has noted that the *matK* gene is recognized for its substantial variability among cp genes, attributed to elevated substitution rates in all codon positions [29]. Nevertheless, phylogenetic tree analyses revealed that the *H. obtusifolia* (NC_044833), *H. aspera* (MG524992) and *H. davidii* (NC_050783) species were not clustered with their respective species/clades. Instead, *H. aspera* and *H. davidii* (NC_050783) individuals were clustered with *H. strigosa* with a strong bootstrap value (84%) (Supplementary Fig. S24). Furthermore, the sequence similarity between *H. aspera* (excluding MG524992), *H. davidii* (NC_050783) and *H. strigosa* was 100%. Similarly, the other *H. davidii* species were clustered with *H. ampla*. Correspondingly, *H. obtusifolia* (OL537854) was clustered with the *H. obtusifolia* species used in this study and other species of the sect. *Decumaria* with a strong bootstrap value (0.95). In contrast, *H. obtusifolia* (NC_044833) utilized by [18] clustered with sect. *Schizophragma* rather than sect. *Decumaria* (Supplementary Fig. S24). According to these results, we suggest that the species *H. obtusifolia* (NC_044833), *H. aspera* (MG524992) and *H. davidii* (NC_050783) could have been misidentified by [18]. It is noteworthy that the plant flower and leaf morphology of *H. obtusifolia* vs. *H. hydrangeoides* and of *H. aspera* vs. *H. davidii* are very similar. To support our analyses, the phylogenomic studies conducted by [84] revealed that *H. aspera* clustered wit sect. *Asperae* (Clade I), while *H. davidii* clustered with sect. *Chinenses* (Clade II). However, in contrast to [84] phylogenomic studies, *H. obtusifolia* (sect. *Decumaria*) exhibited a paraphyletic relationship. In our studies, all species belonging to sect. *Decumaria* formed a monophyletic group with strong BS value. Therefore, more individuals are required to understand the position of these three species in the Hydrangeeae clade.

In our current investigation, our primary focus was on *H. petiolaris*, a species with a restricted distribution in the southern regions of the Korean peninsula, encompassing Jeju and Ulleung Islands, as well as Japan and Sakhalin, Russia [9,36]. We collected three samples from Jeju Island: one from the 1100-meter highland of Mt. Halla (JJCH), another from Aewol (JJAW; 714 m), and a third from Saryeoni-Supgil (JJSR; 520 m). Additionally, we obtained one sample from Ulleung Island (UL01; 940 m). Our analysis compared these samples with a previously reported *H. petiolaris* individual from Ulleung Island (NC_034936). Our phylogenomic analyses revealed interesting patterns. The *H. petiolaris* sample from Mt. Halla (JJCH) showed a closer genetic affinity to the Ulleung Island populations than to the other Jeju Island populations (JJAW and JJSR). However, the support for this relationship, as indicated by the bootstrap value, was relatively low (Fig. 5; Supplementary Figs. S12-S13). On the other hand, *H. petiolaris* individuals from the remaining Jeju Island populations (JJAW and JJSR) formed a distinct clade. In contrast, the analysis of the entire cp genome suggested that UL01 from Ulleung Island was clustered with Mt. Halla (JJCH) on Jeju Island and another individual from Ulleung Island (Supplementary Fig. S11). However, the bootstrap value for the relationship between UL01 and JJCH was only 15%. As a result, our phylogenetic analysis, including PCS, AA, and Bayesian analyses, suggested that Mt. Halla populations might have dispersed to Ulleung Island; it is essential to note that the data do not strongly support the evidence for this hypothesis.

To enhance our understanding of the molecular age of these

populations, we estimated the divergence times between the Ulleung Island (UL01) and Mt. Halla, Jeju Island (JJCH) populations of *H. petiolaris*, as well as between the populations on Jeju Island (JJAW and JJSR). The separation between Ulleung Island and Mt. Halla took place during the Pleistocene epoch and was estimated to have occurred approximately 0.27 million years ago (Ma), with a 95% highest posterior density (HPD) ranging from 0.08 to 0.65 Ma. The emergence of the crown nodes for UL01 and JJCH was estimated at approximately 0.13 Ma (95% HPD = 0.03–0.31 Ma). Likewise, the crown nodes for JJSR and JJAW were estimated to have originated approximately 30,000 years ago (95% HPD = 0–0.16 Ma) (Fig. 6**A**). Additionally, we performed in-depth analyses using biogeographical methods (BGB, S-DIVA, BBM, and DEC), which suggested that *H. petiolaris* from Mt. Halla (JJCH) likely dispersed relatively recently, approximately 0.13 Ma, followed by JJSR and JJAW approximately 30,000 years ago. However, despite the recent dispersal, the origin of *H. petiolaris* remains uncertain due to the absence of populations from Japan and Sakhalin, Russia, in the present study.

In our earlier studies, we analyzed cp and ITS regions from several *H. petiolaris* individuals from Jeju and Ulleung Islands and Sapporo, Hokkaido, Japan [36]. The results showed two distinct clades: one comprising Mt. Halla, Ulleung, and Japan populations and the other comprising the Aewol and Saekdal populations of Jeju Island. These findings support our present hypothesis that *H. petiolaris* may have dispersed from Mt. Halla, Jeju. Previous studies have speculated on the migration of plants from the Korean peninsula to Japan through Ulleung Island during the glacial era (<2 Ma) [37]. Japan's separation from the Korean Peninsula occurred in the Pleistocene era (1.6–0.01 Ma) [79], while Ulleung Island's volcanic formation took place at approximately 2.7–0.01 Ma [35]. Jeju Island emerged relatively recently from the Korean Peninsula during the 0.01–1.2 Ma period [33]. Seed dispersal from the Korean peninsula to the Japanese archipelago, including Ulleung and Dokdo islands, is likely facilitated by prevailing winds, water and animals [10,67]. Ocean currents, such as the East Korean warm and Tsushima currents, also contribute to potential seed dispersal from the Korean peninsula to Ulleung Island and Japan [42]. However, based on the phylogenomic, molecular clock and biogeographical analyses, we could not definitively conclude the origin of *H. petiolaris* in the present study. Given these results, our forthcoming research endeavors will encompass additional populations from the islands of Jeju and Ulleung, as well as Japan and Sakhalin, Russia. This approach aims to provide more comprehensive insight into the evolutionary origin patterns of *H. petiolaris*.

Similar to *H. petiolaris*, *H. luteovenosa* is distributed on Jeju Island, Korea and Japan and diversified at approximately 2 Ma (Pleistocene Epoch). Our hypothesis suggests that the propagules of Hydrangeeae are small in size and could have been dispersed by wind, water, or animals. Throughout the Pleistocene era, climate played a pivotal role in impacting the level of the East Sea due to the cyclic glacial and interglacial periods, leading to fluctuations and alterations in sea levels [26]. These dynamic environmental shifts likely facilitated the formation of dispersal corridors between Jeju Island, Korea and the Japanese Islands, making gene flow and migration among these regions possible for Hydrangeeae.

The analysis of divergence times indicated that the initial divergence of the Hydrangeaceae family stem lineage was estimated to have occurred approximately 59.49 million years ago during the early Paleocene (95% HPD 48.33–71.9 Ma) (Fig. 6). The phylogenetic informativeness profile for the *ycf1* gene peaks at 73.25, a point that occurs subsequent to the divergence of the tribe Hydrangeeae tribe at time 27.57, and other genes, such as *ndhF*, *matK*, *rpoC2* and *rpoB,* reach higher informativeness (Fig. 6**B**). The remaining plastid genes exhibit shallower, more gradual profiles, steadily gaining informativeness as they extend toward more ancient nodes in the phylogenetic tree [13].

Our biogeographical analyses, employing various ancestral area reconstruction methods, such as S-DIVA, DEC, S-DEC, BBM, BioGeoBEARS (BGB), and BayArea (BA), have provided valuable insights into the evolutionary history and origin of the Hydrangeaceae family, particularly the tribe Hydrangeeae (Fig. 7**;** Supplementary Figs. S14-S18). Through this comprehensive set of analyses, we explored the potential ancestral areas and dispersal patterns of these plant groups. Consistent findings from DEC, S-DEC, and BGB analyses suggest that the potential areas of origin of the Hydrangeaceae family were in North America, whereas the origin for Loasaceae was in South America. This indicates that the diversification of the Hydrangeaceae family likely began in North America, with subsequent events of vicariance and dispersal leading to its expansion into other regions, including Asia and Africa. This is in contrast to the BBM and S-DIVA analyses that present different scenarios regarding the origin of the Hydrangeaceae family. BBM suggests a possible origin in North America and a combined area of Asia & North America, while S-DIVA indicates that Asia was the likely place of origin for the Hydrangeaceae family. To strengthen our analysis, we considered previous studies that also suggested that the origin of the Hydrangeaceae family was North America and Asia [18].

Regarding the tribe Hydrangeeae, both S-DIVA and BBM consistently point to East Asia as the origin of the *Hydrangea* genus (Supplementary Figs. S14, S17, S19, and S22). However, the DEC, S-DEC, and BGB analyses suggest that *Hydrangea* originated in East Asia and North America (Figs. 7 and 8**,** Supplementary Figs. S15, S16, S20, S21, and S23). For instance, the species *H. arborescens*, *H. barbara*, and *H. quercifolia* are distributed in eastern North America, while *H. arguta* is restricted to an isolated geographical location, being the only member of the Hydrangeeae tribe endemic to the Hawaiian Islands [13]. The remaining Hydrangeeae tribes are distributed on the Asian continent. This may arise from the complexities of the biogeographical processes involved in Hydrangeeae tribe evolution. Hence, drawing upon our findings, we hypothesize that the tribe Hydrangeeae most likely had its origins in East Asia and North America during the middle Oligocene epoch. Subsequently, this group experienced substantial diversification, primarily within East Asia. An earlier study also proposed that climate fluctuations among continents during the Tertiary period, coupled with geographic dispersals followed by isolation and habitat distinctions, probably played a pivotal role in instigating the divergence of the Hydrangeaceae family. This divergence, in turn, led to variations in characteristics such as habit/ecology, morphology, and geographic distribution, as described by [87]. In the contemporary era, the majority of Hydrangeaceae members are characterized as temperate herbs or shrubs, primarily found in northern regions. In contrast, Loasaceae members are typically herbaceous plants thriving in xeric regions of the Americas [18].

The Cretaceous period played a pivotal role in the emergence and proliferation of major angiosperm lineages, as more than half of existing plant families can trace their origins back to this era, notably the mid- and late Cretaceous periods [15,4,49,69,81]. Our data support the notion that the Hydrangeaceae and Loasaceae families also originated in the late Cretaceous period. In this epoch, paleontological evidence suggests that Eurasia, North America, and Africa were in closer proximity, facilitating the possibility of migration between these regions [20]. The eventual fragmentation of the supercontinent Laurasia most likely took place during the Paleocene epoch. Following this event, migration between Eurasia and North America persisted via the Bering Land Bridge (BLB) and the North Atlantic Bridge (NALB) [21,22,54,75].

Throughout the Paleocene to the Miocene epochs, the BLB served as a connecting link between Asia and North America, facilitating the expansion of temperate plant taxa [21,75]. Simultaneously, since the early Paleocene, the NALB extended from North America and Greenland to northeastern Europe, primarily serving as the pathway for tropical plant taxa migration [76,77]. Consequently, it is reasonable to speculate that the BLB played a pivotal role in enabling intercontinental floristic exchanges involving Hydrangeeae during the early Miocene epoch. The initial dispersal event between East Asia and North America, followed by the divergence of *H. arguta* in the late Oligocene or early Miocene epoch,

likely transpired through a combination of long-distance dispersal and vicariance, eventually leading to colonization on the Hawaiian Islands. Importantly, previous studies have indicated that the BLB, or potentially some disjunct islands, may have served as a viable conduit for species dispersal, possibly extending into the Pliocene epoch or beyond [22]. Even in the absence of a continuous land connection between North and South America until the formation of the Isthmus of Panama approximately 3 million years ago, it is probable that plant migration between these continents occurred over the course of the past 50 million years [14]. Our DEC, S-DEC, and BGB reconstruction (Figs. 7 and 8**,** Supplementary Figs. S15, S16, S20, S21, and S23**)** support the notion that East Asia and North America constitute the ancestral area of the Hydrangeeae tribe. Therefore, we propose that the tribe Hydrangeeae most likely originated in East Asia and North America, subsequently diversifying into East Asia via the BLB during the Oligocene period.

In summary, our study makes a substantial contribution to elucidating the evolutionary history of the Hydrangeaceae family. The newly assembled plastid genomes and the phylogenomic analysis offer valuable insights into the systematic relationships within the Hydrangeeae tribe. Nevertheless, to enhance our understanding of the phylogenetics and biogeography of *H. petiolaris* and other Hydrangeaceae family members, further research involving a broader sampling of populations and increased molecular data is warranted.

5. Conclusion

Our study presented 28 newly assembled plastomes belonging to the Hydrangeaceae family. Furthermore, we established an evolutionary framework aimed at assessing the systematic and biographical history of the Hydrangeeae tribe. Distinct from previous research, which mainly focused on the phylogenetic relationships within the Cornales order, this study represents the first phylogenomic analysis specifically targeting the Hydrangeeae tribe. We observed a significant degree of conservation in both gene content and arrangement across all the plastid genomes. Nonetheless, we also detected supplementary 18 bp tandem repeats within the *Macrophyllae* and *Chinenses* sections, with four copies present in the *Dichroa* section, except for *H. febrifuga*, which exhibited two copies. Further structural validation and qPCR analyses revealed the likely functional role of *accD* with lower expression. The comprehensive phylogenetic analyses, encompassing the whole cp genome, concatenated 79 protein-coding genes, translated amino acids, and Bayesian datasets, provided strong support for the polyphyletic nature of the Hydrangeeae tribe. This finding led to its division into two distinct sections, supported by high bootstrap values. Furthermore, we explored the phylogenetic relationships of *H. petiolaris* between the Jeju and Ulleung Island populations. Our findings suggest the need for additional studies incorporating more samples and molecular data to gain a comprehensive understanding of the phylogenomics of *H. petiolaris*. Through the estimation of divergence times and the implementation of biogeographical analyses, our study suggests that the common ancestors of the Hydrangeeae tribe probably originated from North America and East Asia, experiencing diversification primarily within East Asia during the Paleocene epoch. Consequently, we posit East Asia and North America as the principal centers of origin for taxa found in these regions. We hypothesize that the Bering Land Bridge played a crucial role in facilitating the migration of the Hydrangeeae tribe between East Asia and North America.

Ethics approval and consent to participate

Our study did not require ethics approval or consent for participation.

Consent to publish

All authors reviewed and endorsed the final manuscript.

Funding

Funding from the National Research Foundation of Korea (NRF) (2021R1A2C3010477), Ministry of Education, the Republic of Korea, was provided to SeonJoo Park, Department of Life Sciences, Yeungnam University.

CRediT authorship contribution statement

GR, SJP and TJY conceived the study; KC, EML, CWM, HS and JSK collected samples; GR, KC and EML assembled the genome; GR annotated the genome, designed and performed the experiments, conducted all data analyses, prepared figures, and wrote the manuscript; SJP supervised the project. All the authors have read and approved the final manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

All the data presented in this article are accurate and reliable.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.10.010.

References

[1] Ahmed I, Biggs PJ, Matthews PJ, Collins LJ, Hendy MD, Lockhart PJ. Mutational dynamics of aroid chloroplast genomes. Genome Biol Evol 2012;4:1316–23.
[2] Asaf S, Waqas M, Khan AL, Khan MA, Kang SM, Imran QM, Shahzad R, Bilal S, Yun BW, Lee IJ. The Complete Chloroplast Genome of Wild Rice (Oryza minuta) and Its Comparison to Related Species. Front Plant Sci 2017;8:304.
[3] Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, Wishart DS. Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res 2016;44: W147–53.
[4] Barba-Montoya J, Dos Reis M, Schneider H, Donoghue PCJ, Yang Z. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a Cretaceous Terrestrial Revolution. N Phytol 2018;218:819–34.
[5] Barrett CF, Baker WJ, Comer JR, Conran JG, Lahmeyer SC, Leebens-Mack JH, Li J, Lim GS, Mayfield-Jones DR, Perez L, Medina J, Pires JC, Santos C, Stevenson DW, Zomlefer WB, Davis JI. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. N Phytol 2016;209:855–70.
[6] Bouckaert R, Vaughan TG, Barido-Sottani J, Duchene S, Fourment M, Gavryushkina A, Heled J, Jones G, Kuhnert D, De Maio N, Matschiner M, Mendes FK, Muller NF, Ogilvie HA, du Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C, Stadler T, Drummond AJ. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol 2019;15:e1006650.
[7] Chen JH, Hao ZD, Xu HB, Yang LM, Liu GX, Sheng Y, Zheng C, Zheng WW, Cheng TL, Shi JS. The complete chloroplast genome sequence of the relict woody plant Metasequoia glyptostroboides Hu et Cheng. Front Plant Sci 2015;6.
[8] Chen VB, Arendall 3rd WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 2010;66:12–21.
[9] Cho JS, Jeong JH, Kim SY, Lee JY, Lee CH. Several factors affecting seed germination of *Hydrangea petiolaris* Siebold & Zucc. Korean J Plant Resour 2014;27: 534–9.
[10] Cho MS, Takayama K, Yang J, Maki M, Kim SC. Genome-wide single nucleotide polymorphism analysis elucidates the evolution of prunus takesimensis in ulleung island: the genetic consequences of anagenetic speciation. Front Plant Sci 2021;12: 706195.
[11] Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK. The complete chloroplast genome sequence of Pelargonium x hortorum: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol 2006;23:2175–90.
[12] Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci: a Publ Protein Soc 1993;2:1511–9.

[13] De Smet Y, Granados Mendoza C, Wanke S, Goetghebeur P, Samain MS. Molecular phylogenetics and new (infra)generic classification to alleviate polyphyly in tribe Hydrangeeae (Cornales: Hydrangeaceae). Taxon 2015;64:741–53.
[14] Domingo L, Tomassini RL, Montalvo CI, Sanz-Perez D, Alberdi MT. The Great American Biotic Interchange revisited: a new perspective from the stable isotope record of Argentine Pampas fossil mammals. Sci Rep 2020;10:1608.
[15] Foster CSP, Sauquet H, van der Merwe M, McPherson H, Rossetto M, Ho SYW. Evaluating the Impact of Genomic Data and Priors on Bayesian Estimates of the Angiosperm Evolutionary Timescale. Syst Biol 2017;66:338–51.
[16] Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res 2004;32:W273–9.
[17] Fu CN, Li HT, Milne R, Zhang T, Ma PF, Yang J, Li DZ, Gao LM. Comparative analyses of plastid genomes from fourteen Cornales species: inferences for phylogenetic relationships and genome evolution. Bmc Genom 2017;18.
[18] Fu CN, Mo ZQ, Yang JB, Ge XJ, Li DZ, Xiang QY, Gao LM. Plastid phylogenomics and biogeographic analysis support a trans-Tethyan origin and rapid early radiation of Cornales in the Mid-Cretaceous. Mol Phylogenetics Evol 2019;140.
[19] Gao C, Li T, Zhao X, Wu C, Zhang Q, Zhao X, Wu M, Lian Y, Li Z. Comparative analysis of the chloroplast genomes of Rosa species and RNA editing analysis. BMC Plant Biol 2023;23:318.
[20] Gheerbrant E, Rage JC. Paleobiogeography of Africa: How distinct from Gondwana and Laurasia? Palaeogeogr Palaeoclimatol Palaeoecol 2006;241:224–46.
[21] Gladenkov AY, Oleinik AE, Marincovich L, Barinov KB. A refined age for the earliest opening of Bering Strait. Palaeogeogr Palaeoclimatol Palaeoecol 2002;183:321–8.
[22] Graham A. The role of land bridges, ancient environments, and migrations in the assembly of the North American flora. J Syst Evol 2018;56:405–29.
[23] Granados Mendoza C, Wanke S, Salomo K, Goetghebeur P, Samain MS. Application of the phylogenetic informativeness method to chloroplast markers: A test case of closely related species in tribe Hydrangeeae (Hydrangeaceae). Mol Phylogenetics Evol 2013;66:233–42.
[24] Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res 2019;47:W59–64.
[25] Gurdon C, Maliga P. Two distinct plastid genome configurations and unprecedented intraspecies length variation in the accD coding region in Medicago truncatula. DNA Res 2014;21:417–27.
[26] Harrison SP, Yu G, Takahara H, Prentice IC. Palaeovegetation. Diversity of temperate plants in east Asia. Nature 2001;413:129–30.
[27] He LF, Qiang SJ, Zhang YH, Li HF. The complete chloroplast genome of Hydrangea davidii (Hydrangeaceae). Mitochondrial DNA B Resour 2020;5:3605–7.
[28] Hufford L. Ontogeny and morphology of the fertile flowers of Hydrangea and allied genera of tribe Hydrangeeae (Hydrangeaceae). Bot J Linn Soc 2001;137:139–87.
[29] Hufford L, Moody ML, Soltis DE. A phylogenetic analysis of hydrangeaceae based on sequences of the plastid gene matk and their combination with rbcl and morphological data. Int J Plant Sci 2001;162:835–46.
[30] Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Muller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, Boore JL. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci USA 2007;104:19369–74.
[31] Jo YD, Park J, Kim J, Song W, Hur CG, Lee YH, Kang BC. Complete sequencing and comparative analyses of the pepper (Capsicum annuum L.) plastome revealed high frequency of tandem repeats and large insertion/deletions on pepper plastome. Plant Cell Rep 2011;30:217–29.
[32] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596:583–9.
[33] Kang H-G, Kim C-S, Kim E-S. Human influence, regeneration, and conservation of the Gotjawal forests in Jeju Island, Korea. J Mar Isl Cult 2013;2:85–92.
[34] Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 2002;30:3059–66.
[35] Kim B, Lee Y, Koh B, Jhang SY, Lee CH, Kim S, Chi WJ, Cho S, Kim H, Yu J. Distinctive origin and evolution of endemic thistle of Korean volcanic island: Structural organization and phylogenetic relationships with complete chloroplast genome. PLoS One 2023;18:e0277471.
[36] Kim H, Park KT, Park S. Molecular phylogenetic study of Korean *Hydrangea* L. J Korean Soc Bot Resour 2016;29:407–18.
[37] Kong W-s, Lee S-g, Park H-n, Lee Y-m, Oh S-h. Time-spatial distribution of Pinus in the Korean Peninsula. Quat Int 2014;344:43–52.
[38] Kuang DY, Wu H, Wang YL, Gao LM, Zhang SZ, Lu L. Complete chloroplast genome sequence of Magnolia kwangsiensis (Magnoliaceae): implication for DNA barcoding and population genetics. Genome 2011;54:663–73.
[39] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–9.
[40] Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–86.
[41] Lee HO, Joh HJ, Kim K, Lee SC, Kim NH, Park JY, Park HS, Park MS, Kim S, Kwak M, Kim KY, Lee WK, Yang TJ. Dynamic Chloroplast Genome Rearrangement and DNA Barcoding for Three Apiaceae Species Known as the Medicinal Herb "Bang-Poong". Int J Mol Sci 2019;20.
[42] Lee J-H, Lee D-H, Choi B-H. Phylogeography and genetic diversity of East Asian Neolitsea sericea (Lauraceae) based on variations in chloroplast DNA sequences. J Plant Res 2013;126:193–202.
[43] Lee J, Lee SC, Joh HJ, Lee H, Sung SH, Kang JH, Lee TJ, Yang TJ. The complete chloroplast genome sequence of a Korean indigenous ornamental plant Hydrangea serrata for. fertilis Nakai (Hydrangeaceae). Mitochondrial DNA Part B-Resour 2016;1:27–8.
[44] Lei W, Ni D, Wang Y, Shao J, Wang X, Yang D, Wang J, Chen H, Liu C. Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of Astragalus membranaceus. Sci Rep 2016;6:21669.
[45] Lenz H, Knoop V. PREPACT 2.0: Predicting C-to-U and U-to-C RNA Editing in Organelle Genome Sequences with Multiple References and Curated RNA Editing Annotation. Bioinform Biol Insights 2013;7:1–19.
[46] Li HT, Yi TS, Gao LM, Ma PF, Zhang T, Yang JB, Gitzendanner MA, Fritsch PW, Cai J, Luo Y, Wang H, Van der Bank M, Zhang SD, Wang QF, Wang J, Zhang ZR, Fu CN, Yang J, Hollingsworth PM, Chase MW, Soltis DE, Soltis PS, Li DZ. Origin of angiosperms and the puzzle of the Jurassic gap. Nat Plants 2019;5:461–70.
[47] Li J, Su Y, Wang T. The Repeat Sequences and Elevated Substitution Rates of the Chloroplast accD Gene in Cupressophytes. Front Plant Sci 2018;9:533.
[48] Lopez-Giraldez F, Townsend JP. PhyDesign: an online application for profiling phylogenetic informativeness. BMC Evol Biol 2011;11:152.
[49] Magallon S, Gomez-Acevedo S, Sanchez-Reyes LL, Hernandez-Hernandez T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. N Phytol 2015;207:437–53.
[50] Magee AM, Aspinall S, Rice DW, Cusack BP, Semon M, Perry AS, Stefanovic S, Milbourne D, Barth S, Palmer JD, Gray JC, Kavanagh TA, Wolfe KH. Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Res 2010;20:1700–10.
[51] McDonald MJ, Wang WC, Huang HD, Leu JY. Clusters of nucleotide substitutions and insertion/deletion mutations are associated with repeat sequences. PLoS Biol 2011;9:e1000622.
[52] McManus HA, Fucikova K, Lewis PO, Lewis LA, Karol KG. Organellar phylogenomics inform systematics in the green algal family Hydrodictyaceae (Chlorophyceae) and provide clues to the complex evolutionary history of plastid genomes in the green algal tree of life. Am J Bot 2018;105:315–29.
[53] Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Giray JC, Morden CW, Calie PJ, Jermiin LS, Wolfe KH. Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. Plant Cell 2001;13:645–58.
[54] Morley RJ. Interplate dispersal paths for megathermal angiosperms. Perspect Plant Ecol Evol Syst 2003;6:5–20.
[55] Munoz-Rodriguez P, Carruthers T, Wood JRI, Williams BRM, Weitemier K, Kronmiller B, Ellis D, Anglin NL, Longway L, Harris SA, Rausher MD, Kelly S, Liston A, Scotland RW. Reconciling Conflicting Phylogenies in the Origin of Sweet Potato and Dispersal to Polynesia. Curr Biol 2018;28:1246 (-+).
[56] Nock CJ, Waters DLE, Edwards MA, Bowen SG, Rice N, Cordeiro GM, Henry RJ. Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol J 2011;9:328–33.
[57] Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics 2004;20:289–90.
[58] Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics 2005;21:676–9.
[59] Raghuraman P, Jesu Jaya Sudan R, Lesitha Jeeva Kumari J, Sudandiradoss C. Systematic prioritization of functional hotspot in RIG-1 domains using pattern based conventional molecular dynamic simulation. Life Sci 2017;184:58–70.
[60] Raman G, Nam GH, Park S. Extensive reorganization of the chloroplast genome of Corydalis platycarpa: A comparative analysis of their organization and evolution with other Corydalis plastomes. Front Plant Sci 2022;13:1043740.
[61] Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol 2018;67:901–4.
[62] Ree RH, Smith SA. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Syst Biol 2008;57:4–14.
[63] Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 2012;61:539–42.
[64] Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol 2017;34:3299–302.
[65] Saina JK, Li ZZ, Gichira AW, Liao YY. The Complete Chloroplast Genome Sequence of Tree of Heaven (Ailanthus altissima (Mill.) (Sapindales: Simaroubaceae), an Important Pantropical Tree. Int J Mol Sci 2018;19.
[66] Samain MS, Wanke S, Goetghebeur P. Unraveling Extensive Paraphyly in the Genus Hydrangea s. l. with Implications for the Systematics of Tribe Hydrangeeae. Syst Bot 2010;35:593–600.
[67] Seo H-S, Kim S-H, Kim S-C. Chloroplast DNA insights into the phylogenetic position and anagenetic speciation of Phedimus takesimensis (Crassulaceae) on Ulleung and Dokdo Islands, Korea. PLOS ONE 2020;15:e0239734.
[68] Shahzadi I, Abdullah, Mehmood F, Ali Z, Ahmed I, Mirza B. Chloroplast genome sequences of Artemisia maritima and Artemisia absinthium: Comparative analyses, mutational hotspots in genus Artemisia and phylogeny in family Asteraceae. Genomics 2020;112:1454–63.
[69] Silvestro D, Cascales-Minana B, Bacon CD, Antonelli A. Revisiting the origin and diversification of vascular plants through a comprehensive Bayesian analysis of the fossil record. N Phytol 2015;207:425–36.

[70] Soltis DE, Xiang Q-Y, Hufford L. Relationships and Evolution of Hydrangeaceae Based on rbcL Sequence Data. Am J Bot 1995;82:504–14.
[71] Song W, Ji C, Chen Z, Cai H, Wu X, Shi C, Wang S. Comparative Analysis the Complete Chloroplast Genomes of Nine Musa Species: Genomic Features, Comparative Analysis, and Phylogenetic Implications. Front Plant Sci 2022;13:832884.
[72] Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol 2008;57:758–71.
[73] Terrab A, Paun O, Talavera S, Tremetsberger K, Arista M, Stuessy TF. Genetic diversity and population structure in natural populations of Moroccan Atlas cedar (Cedrus atlantica; Pinaceae) determined with cpSSR markers. Am J Bot 2006;93:1274–80.
[74] Tian N, Han LM, Chen C, Wang ZZ. The complete chloroplast genome sequence of Epipremnum aureum and its comparative analysis among eight Araceae species. Plos One 2018;13.
[75] Tiffney BH. THE EOCENE NORTH ATLANTIC LAND BRIDGE: ITS IMPORTANCE IN TERTIARY AND MODERN PHYTOGEOGRAPHY OF THE NORTHERN HEMISPHERE. J Arnold Arbor 1985;66:243–73.
[76] Tiffney BH. Perspectives on the origin of the floristic similarity between Eastern Asia and Eastern North America. J Arnold Arbor 1985;66:73–94.
[77] Tiffney BH, Manchester SR. The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the Northern Hemisphere tertiary. Int J Plant Sci 2001;162:S3–17.
[78] Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res 2017;45:W6–11.
[79] Tominaga O, Su ZH, Kim CG, Okamoto M, Imura Y, Osawa S. Formation of the Japanese Carabina fauna inferred from a phylogenetic tree of mitochondrial ND5 gene sequences (Coleoptera, carabidae). J Mol Evol 2000;50:541–9.
[80] Wallner B, Elofsson A. Can correct protein models be identified? Protein science: a publication of the. Protein Soc 2003;12:1073–86.
[81] Wen J, Nie Z-L, Ickert-Bond SM. Intercontinental disjunctions between eastern Asia and western North America in vascular plants highlight the biogeographic importance of the Bering land bridge from late Cretaceous to Neogene. J Syst Evol 2016;54:469–90.
[82] Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007;35:W407–10.
[83] Yang JB, Yang SX, Li HT, Yang J, Li DZ. Comparative Chloroplast Genomes of Camellia Species. Plos One 2013:8.
[84] Yang X, Zhang X, Xue T, Zhang X, Yang F, Yu J, Janssens SB, Bussmann RW, Yu S. Phylogenomics and historical biogeography of Hydrangeeae (Hydrangeaceae) elucidate the effects of geologic and climatic dynamics on diversification. Proc Biol Sci 2023;290:20230659.
[85] Yang Z, Zhao T, Ma Q, Liang L, Wang G. Comparative Genomics and Phylogenetic Analysis Revealed the Chloroplast Genome Variation and Interspecific Relationships of Corylus (Betulaceae) Species. Front Plant Sci 2018;9:927.
[86] Yu Y, Harris AJ, Blair C, He XJ. RASP (Reconstruct Ancestral State in Phylogenies): A tool for historical biogeography. Mol Phylogenetics Evol 2015;87:46–9.
[87] Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB, Royer DL, Soltis DE, Stevens PF, Westoby M, Wright IJ, Aarssen L, Bertin RI, Calaminus A, Govaerts R, Hemmings F, Leishman MR, Oleksyn J, Soltis PS, Swenson NG, Warman L, Beaulieu JM. Three keys to the radiation of angiosperms into freezing environments. Nature 2014;506:89 (-+).
[88] Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 2008;18:821–9.
[89] Zhang YJ, Du LW, Liu A, Chen JJ, Wu L, Hu WM, Zhang W, Kim K, Lee SC, Yang TJ, Wang Y. The complete chloroplast genome sequences of five epimedium species: lights into phylogenetic and taxonomic analyses. Front Plant Sci 2016;7.