Article

Lineage-Specific Variation in IR Boundary Shift Events, Inversions, and Substitution Rates among Caprifoliaceae *s.l.* (Dipsacales) Plastomes

Seongjun Park [1], **Minji Jun [2]**, **Sunmi Park [2]** and **SeonJoo Park [2,*]**

1   Institute of Natural Science, Yeungnam University, Gyeongsan 38541, Gyeongbuk, Korea;
      seongjun.og@gmail.com
2   Department of Life Sciences, Yeungnam University, Gyeongsan 38541, Gyeongbuk, Korea;
      alswi9180@naver.com (M.J.); psm1128@yu.ac.kr (S.P.)
*   Correspondence: sjpark01@ynu.ac.kr; Tel.: +82-53-810-2377

Abstract: Caprifoliaceae *s.l.* plastid genomes (plastomes) show that one inversion and two inverted repeat boundary shifts occurred in the common ancestor of this family, after which the plastomes are generally conserved. This study reports plastome sequences of five additional species, *Fedia cornucopiae*, *Valeriana fauriei,* and *Valerianella locusta* from the subfamily Valerianoideae, as well as *Dipsacus japonicus* and *Scabiosa comosa* from the subfamily Dipsacoideae. Combined with the published plastomes, these plastomes provide new insights into the structural evolution of plastomes within the family. Moreover, the three plastomes from the subfamily Valerianoideae exhibited accelerated nucleotide substitution rates, particularly at synonymous sites, across the family. The patterns of *accD* sequence divergence in the family are dynamic with structural changes, including interruption of the conserved domain and increases in nonsynonymous substitution rates. In particular, the *Valeriana accD* gene harbors a large insertion of amino acid repeat (AAR) motifs, and intraspecific polymorphism with a variable number of AARs in the *Valeriana accD* gene was detected. We found a correlation between intron losses and increased ratios of nonsynonymous to synonymous substitution rates in the *clpP* gene with intensified positive selection. In addition, two Dipsacoideae plastomes revealed the loss of the plastid-encoded *rps15*, and a potential functional gene transfer to the nucleus was confirmed.

Keywords: *accD*; amino acid repeat motifs; *clpP*; intron loss; positive selection; intracellular gene transfer

1. Introduction

The plastid genome (plastome) of angiosperms is generally conserved and has a quadripartite structure with a pair of inverted repeats (IR) separated by large and small single copy (LSC and SSC) regions [1]. Plastomes generally range from 120 to 160 kb in length and contain 113 unique genes, 79 protein-coding genes, 30 tRNAs, and 4 rRNAs. However, accumulating data on complete plastome sequences exhibit variations in gene and intron content [2]. Plastid gene loss requires functional transfer to the nucleus before the loss in its plastome [3]. Extensive genome rearrangements, including IR boundary shifts, have also been reported for several lineages such as Campanulaceae [4], Caryophyllaceae [5], Fabaceae [6], Geraniaceae [7], Oleaceae [8], and Papaveraceae [9,10].

Lineage-specific variation in the rate of plastome sequence evolution has been documented in angiosperms [5,11–13]. Higher substitution rates are associated with structural rearrangements [5,14]. Accelerated rates of nucleotide substitution are affected by the whole genome as well as a subset of protein-coding genes. Compared with other plastid-encoded genes, acetyl-CoA carboxylase subunit β (*accD*), ATP-dependent Clp protease proteolytic subunit (*clpP*), DNA-directed RNA polymerase subunit α (*rpoA*), some subunits of ribosomal proteins, and the chloroplast factors *ycf1*, *ycf2*, and *ycf4*, show dynamic acceleration [5,9,11,15–18]. Multiple mechanisms, including dysfunction of DNA replication,

repair, and recombination (DNA-RRR) machinery, localized hypermutation, mutagenic retroprocessing, and pseudogenization, have been hypothesized to explain this acceleration [11,14,16,19].

The most extreme accelerations are found in the plastid-encoded *accD* and *clpP*, which have undergone a history of insertions and deletions (indels) of amino acid sequences or intron losses in multiple independent lineages, respectively [9,15]. Several angiosperm lineages have experienced losses in the *accD* or *clpP* genes [10,20,21]. The plastid-encoded *accD* has been functionally replaced via gene transfer to the nucleus or gene substitution in eukaryotic ACCase [15,16,20,22–26]. However, clear evidence of plastid *clpP* transfer to the nucleus in angiosperms is lacking. Highly divergent *accD* and *clpP* genes are likely caused by compensatory mutations in the interactions between nuclear-encoded plastid-targeted subunits [27,28]. This is because the function of the two genes depends on nuclear-encoded proteins that assemble plastid-localized subunits. For example, the plastid-encoded *accD* plays an essential role in the fatty acid biosynthesis pathway [29], which is a subunit of the prokaryotic acetyl–CoA carboxylase (ACC) complex [30]. The prokaryotic ACC complex consists of four subunits, three of which are nuclear-encoded proteins: acetyl–CoA carboxylase subunit α (*ACCA*), biotin carboxyl carrier protein subunit (*ACCB*), and biotin carboxylase subunit (*ACCC*). Plastid-encoded *clpP* is a subunit of the caseinolytic protease (CLP) complex [31] and is involved in multiple processes of chloroplast development [32]. Multiple subunits (CLPP2 to CLPP6) of the CLP complex are encoded in the nucleus [31]. Thus, the nuclear-encoded subunits interact with plastid-encoded subunits.

The honeysuckle family (Caprifoliaceae *sensu lato* [*s.l.*]) comprises approximately 825 species in 28-42 genera with widespread cosmopolitan distribution [33]. Caprifoliaceae *s.l.* has been classified into seven major groups (six subfamilies and one genus): Diervilloideae, Caprifolioideae, Linnaeoideae, Morinoideae, Dipsacoideae, Valerianoideae, and *Zabelia* (Rehder) Makino [34]. A recent phylogenomic study based on nuclear loci and plastome sequences suggested that *Zabelia* is recognized as a new subfamily of Zabelioideae [35]. To date, complete plastomes of 22 genera have been sequenced (National Center for Biotechnology Information; NCBI, accessed on 18 August 2021). The sequenced Caprifoliaceae *s.l.* plastomes range in size from 151.3 to 161.6 kb with a quadripartite organization. Variations in inverted repeat (IR) boundary shifts and nucleotide substitution rates have been documented in the Caprifoliaceae *s.l.* plastomes [34,36]. Despite broader sampling and sequencing across the family, very little is known about the gene evolution, including structural changes and rate variation in the plastid-encoded *accD* and *clpP* genes.

In this study, we generated the complete plastome sequences of three species from the subfamily Valerianoideae and two species from the subfamily Dipsacoideae. Genome organization and nucleotide substitution rates were estimated and compared to the published Caprifoliaceae *s.l.* plastomes. In particular, the correlation between structural evolution and nucleotide substitution rates in the plastid-encoded *accD* and *clpP* genes across this family was examined. In addition, we examined intra- and infraspecific length variations in the *accD* coding region of *Valeriana fauriei* Briq. and *V. sambucifolia* f. *dageletiana* (Nakai ex F.Maek.) Hara.

2. Results

2.1. Plastome Organization

We sequenced and assembled the complete plastomes of three Valerianoideae (*Fedia cornucopiae* (L.) Gaertn., *V. fauriei*, and *Valerianella locusta* (L.) Laterr.) and two Dipsacoideae (*Dipsacus japonicus* Miq. and *Scabiosa comosa* Fisch. ex Roem. & Schult.) species (Figure S1, see Supplementary Materials). Among the five species, the plastome size ranged from 149,809 bp (*V. locusta*) to 160,243 bp (*D. japonicus*) (Table 1). *Dipsacus japonicus* had the largest LSC (87,066 bp), whereas *F. cornucopiae* had the smallest SSC (15,862 bp). The GC content of the *D. japonicus* (38.8%) was higher than that of the other species (Table 1). The three Valerianoideae plastomes encoded 79 protein-coding genes, 30 tRNA genes, and 4 rRNA genes (Table 1). However, the ribosomal protein subunit S15 (*rps15*) appears to

be a pseudogene in two Dipsacoideae plastomes (Table 1). Functional replacement by gene transfer of *rps15* from plastid to the nucleus was detected in *Dipsacus* transcriptome data (Figure S2). The plastome of *D. japonicus* was missing *trnT-GGU* and contained two *trnE-UUC* with 90.4% nucleotide identity (Figure S3). The duplicated gene content in the IR region varied as a result of IR expansion and contraction. The intron content also varied owing to the loss of intron in the *clpP* gene. The plastome of *F. cornucopiae* was missing the first intron of the *clpP* gene. The plastome of *V. locusta* was missing both introns in the *clpP* gene (Figure S1).

Table 1. Comparison of Caprifoliaceae plastomes sequenced in this study.

| Taxon | Dipsacoideae | | | Valerianoideae | |
	Dipsacus japonicus	*Scabiosa comosa*	*Fedia cornucopiae*	*Valeriana fauriei*	*Valerianella locusta*
Size (bp)	160,243	159,651	152,196	155,302	149,809
LSC length (bp)	87,066	87,477	82,960	85,541	82,103
SSC length (bp)	17,850	18,716	15,862	15,159	15,796
IR length (bp)	27,664	26,729	26,687	27,301	25,955
Number of protein-coding genes	78 (6)	78 (7)	79 (4)	79 (4)	79 (4)
Number of tRNA genes	29 (8)	30 (7)	30 (7)	30 (7)	30 (7)
Number of rRNA genes	4 (4)	4 (4)	4 (4)	4 (4)	4 (4)
Number of introns	21 (5)	20 (5)	21 (5)	21 (5)	19 (5)
GC content (%)	38.8	38.7	38.2	38.4	38.1

Among the analyzed Caprifoliaceae *s.l.* plastomes, the largest number of repeat pairs (95) was found in *Morina*, and the fewest repeats (11) were observed in *Valerianella* (Table S1). The average value of repeat pairs from Caprifoliaceae *s.l.* plastomes was ~10 times higher than that from the outgroups (Table S1). The number of repeat pairs from the subfamily Valerianoideae was 2–4 times higher than the average value of repeat pairs from the outgroups, but the number of the repeat pairs from *Valerianella* was similar to that from the outgroups.

To understand the evolutionary history of genome rearrangement in the family, we constructed a phylogenetic tree using 72 plastid genes (Figure 1). The inversions mainly occurred within the IR and SSC regions and some inversion is likely the result of a series of IR expansions and contractions (Figure 1). Based on the most parsimonious interpretation, the ancestral plastome of Caprifoliaceae *s.l.* had three structural changes: (1) an inversion associated with the *ndhF* gene, (2) a contraction at the IR_B/SSC boundary to *trnN* resulting in the entire *ycf1* gene into the SSC region, and (3) a contraction at the IR_A/LSC boundary, from the *rps19* to *rpl23* gene (Figure 1). The plastome rearrangement model suggests that lineage- or species-specific events occurred independently after the ancestral structural changes (Figure 1). For example, an independent contraction to *trnI* at the IR_A/LSC boundary and contraction to *trnR* at IR_A/SSC occurred in early diverging *Weigela*. Mauve alignment among the subfamily Caprifolioideae identified seven locally collinear blocks (LCBs) with two inversions involving eight breakpoints (Figure S4). Within the subfamily Caprifolioideae, a contraction to *ycf2* at IR_A/LSC boundary in *Leycesteria* and a contraction to *trnR* at IR_A/SSC occurred in *Triosteum*, respectively. The inversion in *Leycesteria* is the result of IR expansion at the IR_B/SSC boundary to *ndhH* and IR contraction at the IR_A/SSC boundary to *trnN*, resulting in the relocation of the *ycf1-rps15* region (Figure 1). Mauve alignment among the subfamilies Valerianoideae and Dipsacoideae identified five LCBs with six inversions involving seven breakpoints (Figure S4). Contraction to *trnI* at the IR_A/LSC boundary, followed by an expansion to *trnH* at the IR_B/LSC boundary and an expansion back to *ycf1*, indicated synapomorphic events in the subfamily Valerianoideae (Figure 1). After that, four IR boundary shift events occurred in the common ancestor of the *Fedia*/*Valerianella*/*Valeriana* clade. The *rpl32-ndhF* region in *Fedia*, *Valerianella*, and *Valeriana* plastomes were relocated as a result of two expansions and two contractions.

Expansion at the IR_A/SSC boundary resulted in the duplication of *ccsA* and included a C-terminal portion of *ndhD* (*Fedia*: 1265 bp; *Valerianella*: 1286 bp; *Valeriana*: 1314 bp), generating a truncated *ndhD* fragment in IR_B (Figure 1). Within the subfamily Dipsacoideae, an expansion from *trnH* to *rps3* at the IR_A/SSC boundary and an expansion from *trnN* to *ycf1* occurred in the common ancestor of the *Dipsacus*/*Scabiosa* clade. The inversion event associated with *ccsA-trnL-rpl32-ndhF* is unique to *Scabiosa* (Figure 1). In *Weigela*, two IR contractions occurred at the IR_A/LSC and IR_A/SSC boundaries.
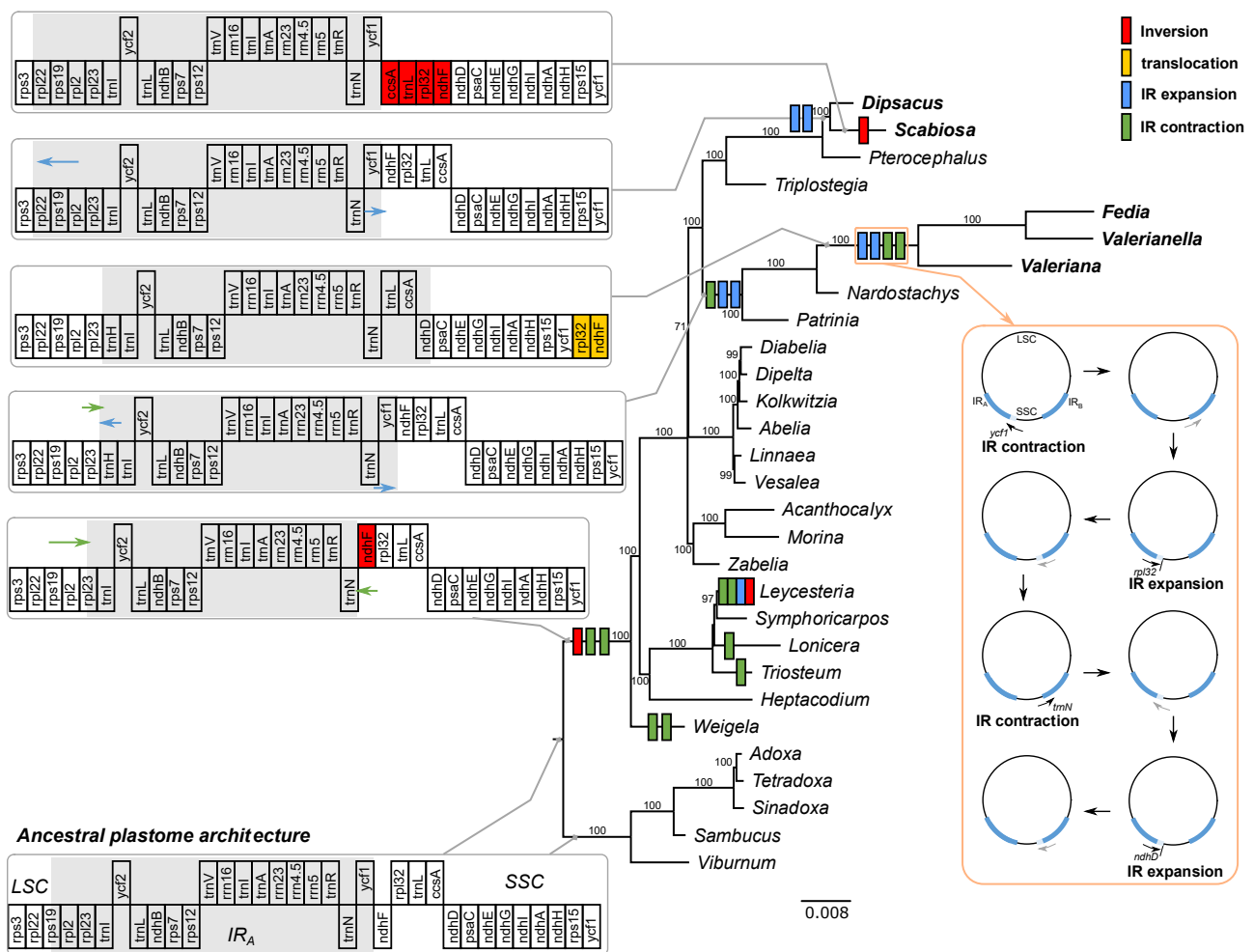

Figure 1. Plastome rearrangement in the analyzed Caprifoliaceae *s.l.* Schematic diagrams (gray open boxes) of the genomic regions surrounding the inverted repeat (IR) region. Genes drawn below the horizontal line indicate sequences found in an inverted orientation. Gray shadings indicate the IR_A region. The hypothetical models for IR expansion and contraction in the common ancestor of Fedia, Valerianella, and Valeriana are illustrated (orange boxes).

2.2. Elevated Substitution Rates in the Plastomes of the Subfamily Valerianoideae

The 24 Caprifoliaceae *s.l.* genera and five Adoxaceae genera shared 72 plastid-encoded genes. We excluded seven protein genes from the concatenated data set and substitution rate analysis because they appeared to be pseudogenes or losses (*clpP*, *rps3*, *rps15*, *ycf1*, and *ycf2*) and divergent (*accD* and *ycf3*) among some species within the family. To examine rate variation in the selected plastid genes among the 24 genera, nonsynonymous (d_N) and synonymous (d_S) substitution rates were estimated using the phylogenetic tree as a constraint tree (Figure 2). The d_N and d_S values in pairwise comparisons between *Viburnum* and Caprifoliaceae *s.l.* showed that *Fedia*, *Valerianella*, and *Valeriana* from the subfamily Valerianoideae had significantly higher d_S rates than the other analyzed species (Wilcoxon rank-sum test, $p < 0.001$, after Bonferroni correction; Table S1), except for the comparison

between *Valeriana* and *Pterocephalus*. *Narodostachys* and *Patrinia* had only significantly higher d_S rates than *Weigela* (Wilcoxon rank-sum test, $p < 0.05$, after Bonferroni correction; Table S2). In the case of d_N rates, only 11 comparisons between *Fedia* and six species (*Kolkwitzia, Symphoricarpos, Triosteum, Triplostegia, Weigela,* and *Zabelia*), between *Valerianella* and four species (*Symphoricarpos, Triplostegia, Weigela,* and *Zabelia*), and between *Valeriana* and *Weigela* showed that *Fedia, Valerianella,* and *Valeriana* had significantly higher rates than the other analyzed species (Wilcoxon rank-sum test, $p < 0.05$, after Bonferroni correction; Table S2).
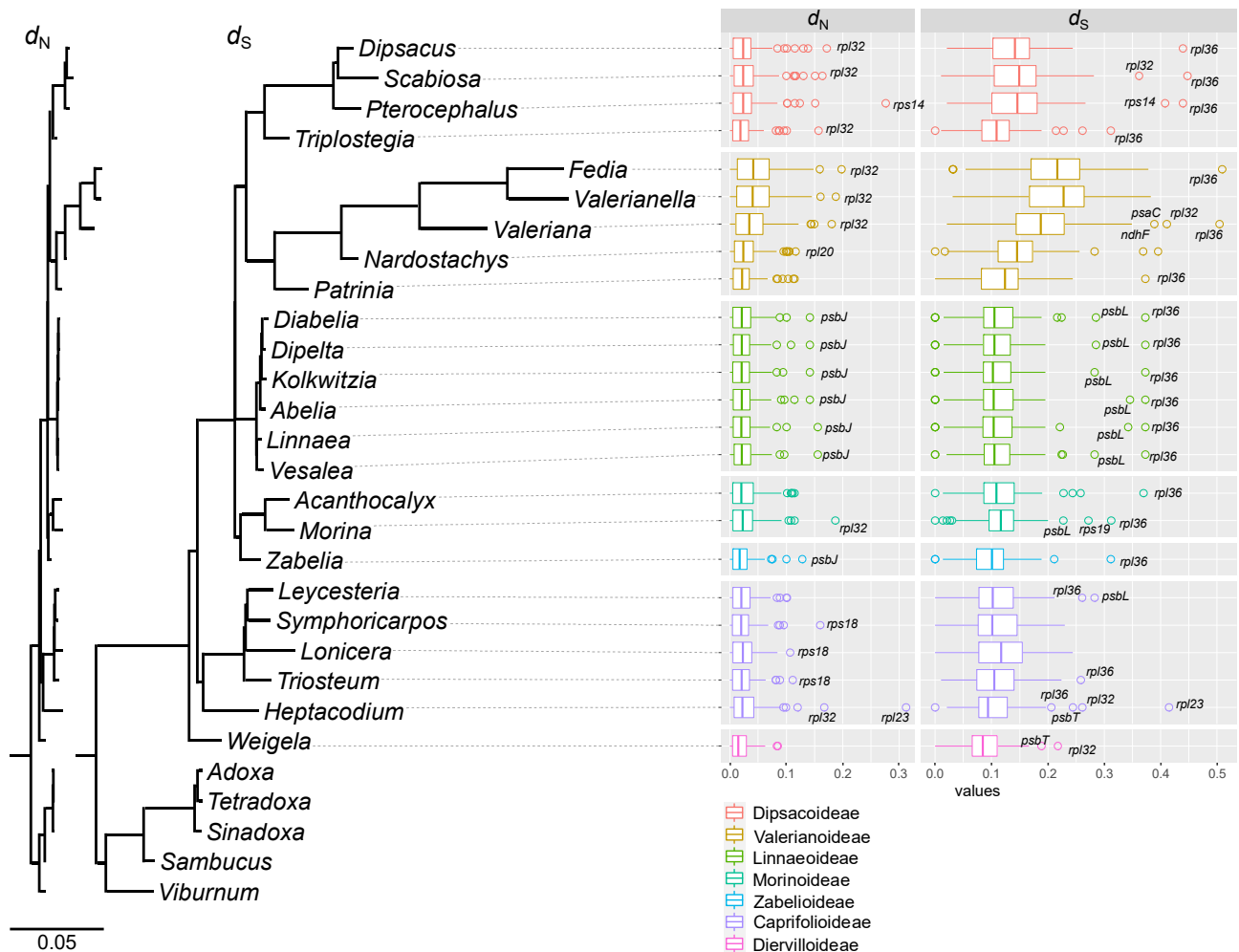

Figure 2. Plastid sequence divergence among the selected Caprifoliaceae *s.l.* Plastid phylograms of nonsynonymous (d_N) and synonymous (d_S) substitution rates based on 72 plastid genes. Boxplots of the values of d_N and d_S for individual genes. The box represents values between quartiles, solid lines extend to the minimum and maximum values, outliers are shown as circles and vertical lines in boxes show median values.

In addition, multiple genes exhibit d_N/d_S ratios that are greater than one, but likelihood ratio tests (LRTs) with Bonferroni correction indicated that d_N/d_S for *infA* in *Acanthocalyx*, *ndhD* in Morinoideae/Linnaeoideae/Valerianoideae/Dipscacoideae/Zabelioideae, *psbM* in *Vesalea*, *rbcL* in *Weigela*, *rpoC1* in *Valeriana*, *rpl22* in *Adoxa/Tetradoxa/Sinadoxa*, *rpl32* in *Heptacodium*, *rps2* in *Acanthocalyx, rps4* in *Scabiosa*, *rps14* in *Pterocephalus*, *rps16* in *Dipsacus/Scabiosa/Pterocephalus*, *rps18* in *Abelia*, and *rps19* in *Acanthocalyx* were significantly different (Table S3).

2.3. Structural Evolution of Plastid-Encoded accD Gene in Caprifoliaceae s.l. Plastomes

The length of the acetyl–CoA carboxylase beta subunit D (*accD*) open reading frame (ORF) varied remarkably among the examined Caprifoliaceae *s.l.* plastomes, ranging from 711 bp in *Heptacodium* to 2517 bp in *Lonicera* (Figure 3). The comparison revealed two histories of insertion events within the N- and C-terminal regions, resulting in the expansion or truncation of the *accD* ORF. The most parsimonious interpretation is that the *accD* genes were interrupted by the insertion of amino acids in the common ancestor of Caprifoliaceae *s.l.* (Figure 3). The second insertion of amino acids occurred in the common ancestor of *Zabelia*, Morinoideae, Linnaeoideae, Valerianoideae, and Dipscacoideae (except *Pterocephalus* and *Triplostegia)*, which was split into two portions of the conserved domain (Figures 3 and S4).
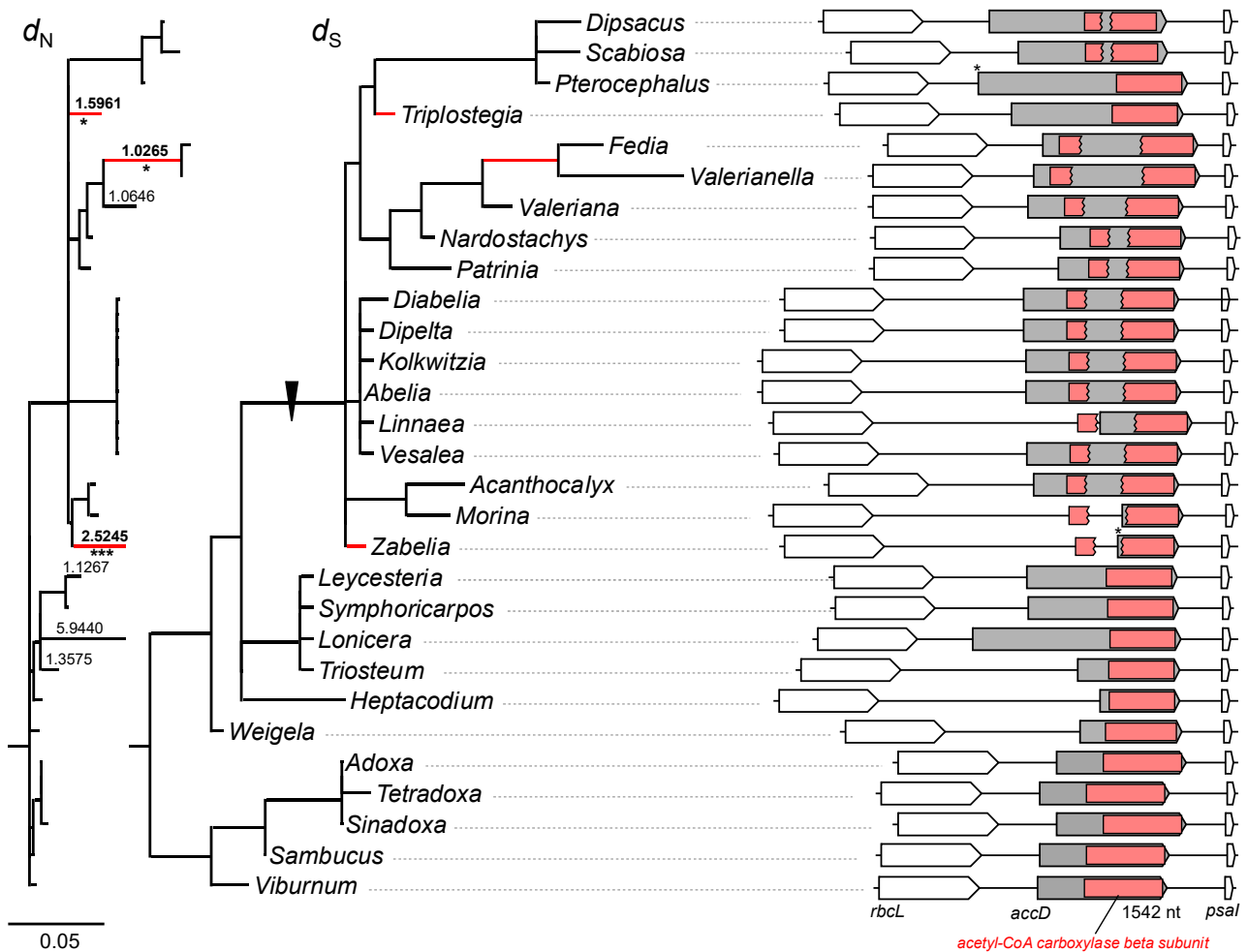

Figure 3. Rapid structural evolution of the *accD* gene. Phylograms show nonsynonymous (d_N) and synonymous (d_S) substitution rates for the *accD* among the analyzed Caprifoliaceae *s.l.* with outgroups. Scale bar indicates the number of substitutions per site. Gray boxes indicate the predicted open reading frames (ORFs). Red rectangles indicate the conserved domain in the ORFs. Asterisks indicate that the ORF has an alternative start codon. Branch lengths are drawn to the same scale based on d_N and d_S substitutions per site. Branches with significantly higher d_N/d_S ratios, as determined by likelihood ratio test are marked with asterisks (*, $p < 0.05$; ***, $p < 0.001$ after Bonferroni correction).

In particular, the *accD* ORFs of *Morina*, *Linnaea*, and *Zabelia* contain a part of the conserved domain, but the catalytic sites are included (Figure S2). Conserved domain (CD) searches identified a portion of Apolipoprotein, MSCRAMM_ClfB, MSCRAMM_SdrC, GAT1, PPK08581, PPK05901, SMC_N, pneumo_PspA, and rplD subfamilies surrounding the conserved domain of *accD* in the predicted ORF (Table S4). The presence of MSCRAMM_ClfB in *Symphoricarpos*, Linnaeoideae, and Adoxaceae indicated that this event occurred in the most recent common ancestor of each clade (Table S4). Protein sequence alignment of all inserted regions from the *accD* was highly divergent with low amino acid identities of 2.9–96.7% (Figure S4). However, two lineages—*Dipsacus/Scabiosa* and Linnaeoideae—had high amino acid identities of 94.3% and 83.1–96.7%, respectively.

The d_N and d_S were calculated for the selected Caprifoliaceae *s.l.* and outgroup using only the conserved domain sequences to test the effect of the insertion on nucleotide substitution rates of the *accD* gene. The d_N values for *accD* in the selected Caprifoliaceae *s.l.* were significantly higher than in the outgroup (Wilcoxon rank-sum test, $p < 0.05$; Figure S3). Seven branches with d_N/d_S ratios > 1 were detected, but LTRs showed that three branches, *Fedia/Valerianella*, *Triplostegia*, and *Zabelia*, were significantly different ($p < 0.00001$ after Bonferroni correction, Figure 3). The RELAX analysis indicated that the *accD* experienced significantly intensified selection in the *Zabelia*, Morinoideae, Linnaeoideae, Valerianoideae, and Dipsacoideae clade ($k = 1.96$, $p = 0.020$, likelihood ratio [LR] = 5.38), and the Caprifoliaceae *s.l.* clade ($k = 2.25$, $p = 0.032$, LR = 4.60).

2.4. Length Variation in the Valeriana accD Gene

Valeriana fauriei plastome contains an expanded *accD* gene, which is interrupted by amino acid repeats (AARs) surrounding the conserved domains. Compared with the other available *Valeriana* plastomes, *V. officinalis* contains a truncated *accD* gene in its genome, whereas *V. sambucifolia* f. *dageletiana* contains an expanded *accD* like *V. fauriei accD* (Figure 4A). To evaluate the variability of the AAR motifs in the *accD* gene of *V. fauriei* and *V. sambucifolia* f. *dageletiana*, we designed a PCR primer that targets two hotspot regions of *accD* (Figure 4A). The amplicon sizes of the region ranged from 966 bp to 1239 bp (Table S5). Alignment of the two region sequences of the 50 individuals with two *accD* sequences from two plastomes revealed intra-and infraspecific variation of the *accD* in *Valeriana* (Figure 4). Two hotspot regions show length variation, consisting of (1) 3 to 10 repeats of "ESTTTESFAQR" and (2) 5 to 14 repeats of "SDSEEDLIKPD", although there are one or three different amino acid sequences (Figure 4).

2.5. Correlation between Structural Change and Substitution Rates in the Plastid-Encoded clpP Gene

The phylogenetic distribution of *clpP* content in the selected Caprifoliaceae plastomes showed that this gene had been pseudogenized multiple times in the family (Figure 5A). BlastN searches using the plastid-encoded *clpP* from *Viburnum* identified partial exons or introns of *clpP* in *Pterocephalus*, *Triplostegia*, *Heptacodium*, *Weigela*, Linnaeoideae, Morinoideae, and Zabeliaoideae, which lacked a conserved domain (Figure 5). Only *Dipsacus*, *Scabiosa*, *Fedia*, *Leycesteria*, *Nardostachys*, *Symphoricarpos*, *Valerianella*, *Valeriana*, *Lonicera*, and *Triosteum* plastomes contain *clpP* genes that differ in intron content (Figure 5A,B). Similar to the *Valerianella clpP* gene, the *clpP* gene of *Leycesteria*, *Lonicera*, *Symphoricarpos*, and *Triosterum* are missing both introns. The phylogenetic distribution indicated that the loss of the second intron occurred in the *Fedia/Valerianella* clade (Figure 5).
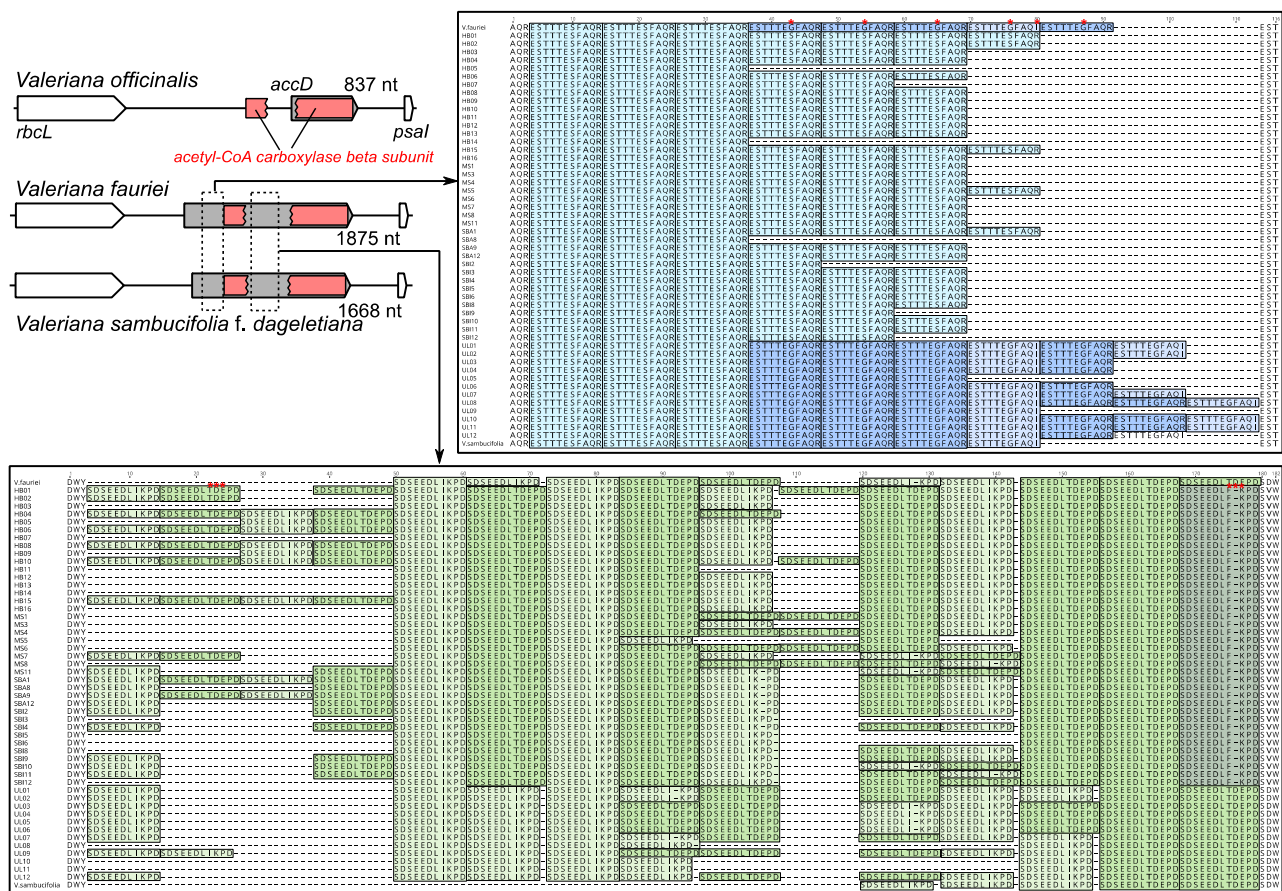
Figure 4. Length variation in the plastid *accD* of *Valeriana*. Schematic diagram of the genomic regions surrounding the *accD* from three *Valeriana* species. Gray boxes indicate the predicted open reading frames (ORFs). Red rectangles indicate the conserved domain in the ORFs. Dashed boxes indicate two hotspot regions in the *accD* gene. Each amino acid sequence of the two hotspot regions of the *accD* copies from 52 *Valeriana* individuals. Blue and green boxes indicate amino acid repeat (AAR) motifs. Asterisks indicate an amino acid sequence mismatch of the AARs (ESTTTESFAQR and SDSEEDLIKPD), respectively. *V. fauriei*: HB, MS, SBA, and SBI; *V. sambucifolia* f. *dageletiana*: UL.

To examine the correlation between structural changes and substitution rates, d_N and d_S were calculated using the 16 *clpP* gene sequences available (Figure 5C). Seven branches with d_N/d_S values >1 were detected. However, LRTs revealed that only four branches, *Fedia*/*Valerianella*/*Valeriana*/*Nardostachys*, Valerianoideae, *Lonicera*, and Caprifolioideae, were significantly under positive selection ($p < 0.00001$ after Bonferroni correction). The RELAX analysis indicated that *clpP* experienced significantly intensified selection in the analyzed Caprifoliaceae *s.l.* ($k = 2.62$, $p = 0.001$, LR = 10.96). Additional analyses indicated that the intronless *clpP* gene experienced significantly intensified selection in the subfamily Caprifolioideae ($k = 3.18$, $p = 0.000$, LR = 16.06). In *Fedia* and *Valerianella*, the *clpP* gene showed intensified selection, but this was not significant ($k = 1.33$, $p = 0.443$, LR = 0.59).
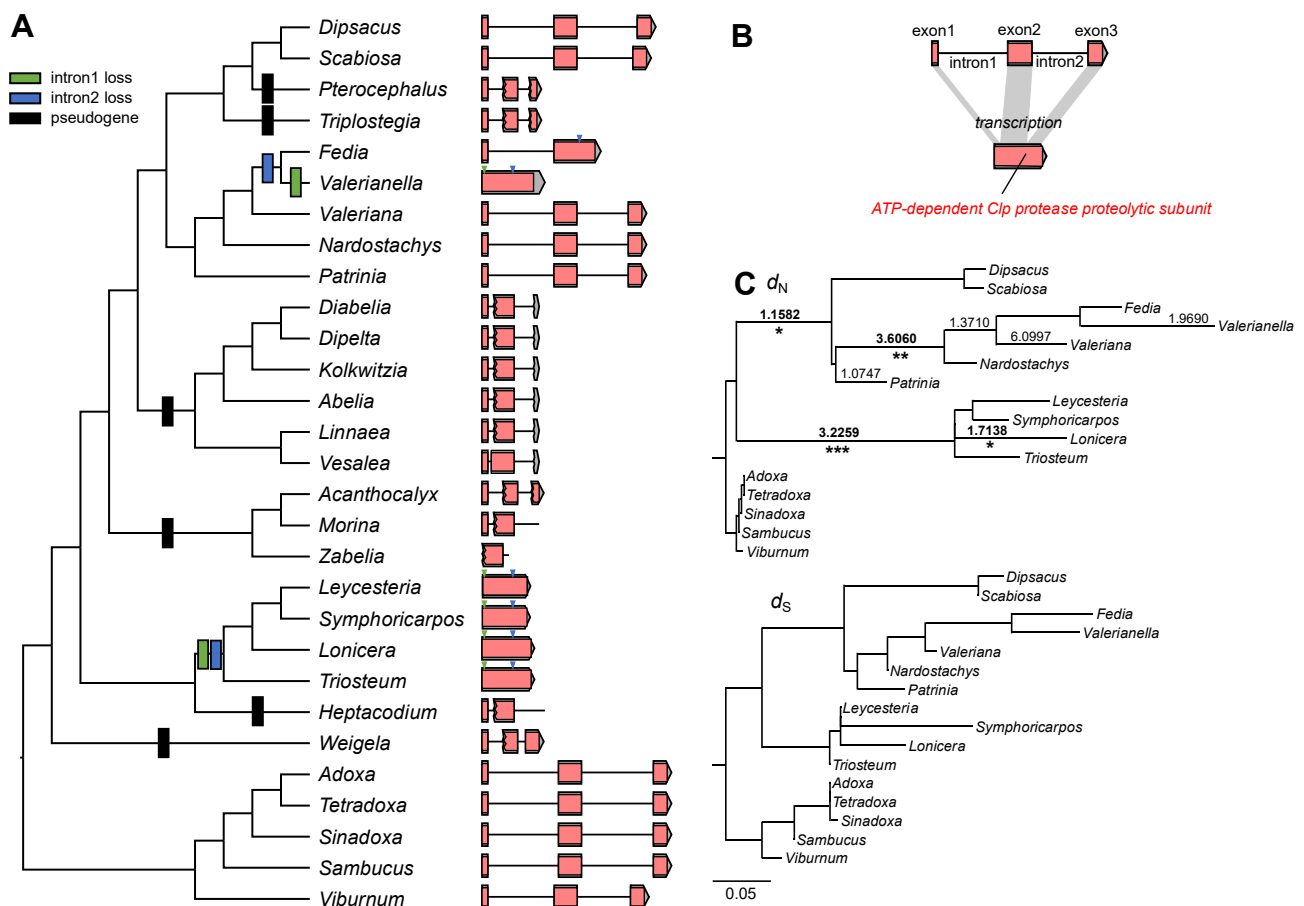
Figure 5. Structural evolution of the *clpP* gene: (**A**) phylogenetic distribution of pseudogene or intron loss among the selected Caprifoliaceae *s.l.* and five outgroups. Arrowheads indicate the positions of the first (green) and second (blue) intron; (**B**) schematic diagram of the structure of *clpP*. Pink boxes indicate the conserved domain of caseinolytic protease; (**C**) phylograms showing nonsynonymous (d_N) and synonymous (d_S) substitution rates for the *clpP* genes among 16 species that have intact gene sequences. Branch lengths are drawn to the same scale based on d_N and d_S substitutions per site. Branches with significantly higher d_N/d_S ratios determined by likelihood ratio test are marked with asterisks (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$ after Bonferroni correction).

3. Discussion

Sequencing of the *F. cornucopiae, V. fauriei, V. locusta, D. japonicus*, and *S. comosa* plastomes revealed that they are distinct from the published plastomes of Caprifoliaceae *s.l.* These plastomes exhibit dynamic changes in structure, gene and intron content, and lineage-specific rate acceleration. Our results also showed a correlation between mutation rates and structural variation in the *accD* and *clpP* genes across Caprifoliaceae *s.l.* and interruption of the *accD* gene in the genus *Valeriana*. The loss of plastid-encoded *rps15* was observed in *D. japonicus* and *S. comosa* plastomes. Comparative analysis of the gene content among Caprifoliaceae *s.l.* plastomes suggests the *rps15* is lost in the common ancestor of *Dipsacus, Scabiosa*, and *Pterocephalus*. Functional replacement of the *rps15* from plastid to the nucleus occurs in these lineages, although we found evidence for a plastid-to-nucleus gene transfer in the *Dipsacus* transcriptome. Additional nuclear transcriptome data for *Scabiosa* and *Pterocephalus* are needed for further investigation.

With the five plastomes, we selected the published plastomes of 19 additional genera from Caprifoliaceae *s.l.* and five genera from Adoxaceae to reconstruct the ancestral plastome in Caprifoliaceae *s.l.* and rearrangement events in each genus. Our results revealed three synapomorphic events (one inversion and two contractions of IR) in the Caprifoliaceae *s.l.* (Figure 1). A previous study showed that the inversion associated

with the *ndhF* gene occurred in Adoxaceae [34]. However, compared with angiosperm plastomes, the ancestral Caprifoliaceae *s.l.* plastome has an inversion between *ndhF*. The comparison of the LCBs and IR boundary shift models indicates that independent events occurred in a lineage- or species-specific manner (Figure 1). The model suggests that the IR expansion and contraction were the main mechanisms for changes in gene order in *F. cornucopiae*, *V. fauriei*, and *V. locusta* plastomes. A double-strand break, followed by strand invasion, expansion, and recombination in IR [37], is a potential mechanism for IR expansion in the family. After ancestral inversion, one additional inversion event associated with *ccsA-trnL-rpl32-ndhF* was identified, unique to *Scabiosa*. Dispersed repeats can cause inversions and there is a correlation between the number of repeats and plastome rearrangements [6,38]. However, *Scabiosa* has a small number of repeats in the family, although it has a relatively large number of repeats compared to the outgroups (Table S1). Moreover, dispersed repeat sequences were not found surrounding the inversion block in the *Scabiosa* plastome, which suggests that a different mechanism of inversion may be involved. Several ebb-and-flow expansions and contractions were also observed in the analyzed Caprifoliaceae *s.l.* plastomes.

Lineage-specific variation occurred in the Caprifoliaceae *s.l.* plastomes, showing that *Fedia, Valeriana*, and *Valerianella* had significantly accelerated d_S in comparison (Figure 2). Structural rearrangements have contributed to higher substitution rates [5,14]. Compared with the analyzed Caprifoliaceae *s.l.* plastomes, the three plastomes showed increased levels of structural divergence (Figure 1). However, IR boundary shifts are the main mechanisms responsible for the genomic changes in this lineage. This variation appears to result from genome-wide acceleration, arguing against localized hypermutation, mutagenic retro-processing, and pseudogenization [16]. Mutated and changed DNA-RRR machinery could be a potential mechanism to explain these phenomena. In the subfamily Valerianoideae, expanded plastome sequencing and examination of organellar-targeted DNA-RRR genes would be needed to explain the causes and consequences of fast-evolving plastomes in this lineage.

Our analysis showed that interrupted *accD* evolution has occurred repeatedly across Caprifoliaceae *s.l.* (Figure 3). Many plastomes show that the *accD* is unrecognized and unannotated because of its extreme divergence [34,36]. In these cases, the *accD* appears to have been lost and possibly transferred to the nucleus, as intracellular gene transfer (IGT) is an ongoing process in angiosperms [3]. Evidence supports the functional replacement of the *accD* gene by gene transfer to the nucleus or gene substitution of nuclear homologs, including coexistence of the nuclear-encoded, plastid-targeted eukaryotic ACCase, prokaryotic ACCase, and the plastid-encoded *accD* [15,16,20,22–26]. However, previous studies have provided some evidence that highly divergent or truncated *accD* genes may be functional in plastids. For example, the divergent copy of the plastid-encoded *accD* in *Lamprocapnos spectabilis* (L.) Fukuhara (Papaveraceae) is transcribed [9]. A functional replacement of truncated *accD* to the nucleus was found in *Trachelium caeruleum* L. (Campanulaceae) [20], *Hypseocharis bilobata* Killip, and *Monsonia emarginata* (L.f.) L'Hér. (Geraniaceae) [15]. In the selected Caprifoliaceae *s.l.* plastomes, many *accD* reading frames remain intact, and several have truncated with the catalytic sites, which suggests that it probably encodes a functional protein. The intensity of both purifying (most branches are $d_N/d_S < 1$) and positive selection (three branches, *Fedia/Valerianella*, *Triplostegia*, and *Zabelia* with $d_N/d_S > 1$) also indicated that the *accD* is under selective constraint or adaptive changes. To fully understand the evolution of ACCase among Caprifoliaceae *s.l.* genomes, searching the assembled nuclear transcriptomes are required. In addition, we examined the variability of the AAR motifs in the *accD* gene and found evidence of intraspecific length variation in the *Valeriana*. One possible mechanism for length polymorphism is replication slippage and recombination [39]. Similar patterns of the *accD* gene have been previously shown to have repetitive amino acid sequence motifs in *L. spectabilis* [9] and *Medicago truncatula* Gaertn. [40]. Gurdon and Maliga [40] suggested that repetitive amino acid motifs within the *accD* could be recombinationally driven.

Previous studies have shown that multiple lineages experience independent loss [7,14,41–43]. Some cases of losses are because the *clpP* gene is unrecognized and unannotated owing to high divergence including structural changes. For example, the *clpP* gene was annotated as a loss in the five Actinidiaceae plastomes [42,43], but the plastomes contain *clpP*-like ORFs that are missing the two introns with a completely conserved domain. The lack of internal stop codons or frameshifts suggests the functionality of the plastid-encoded *clpP*. The *Geranium* and *Monsonia clpP* genes were annotated as pseudogenes or losses [7,14]; however, extremely divergent and intronless *clpP* ORFs are found in the lineages [15]. *ClpP* is involved in important chloroplast processes [32]. If the divergent ORFs are pseudogenes, functional replacement by gene transfer or gene substitution must occur. However, the evolutionary fate of the plastid-encoded *clpP* loss has not been reported in angiosperms. We identified at least six potential pseudogenizations of *clpP* and independent losses of one or both introns in the analyzed Caprifoliaceae *s.l.* (Figure 5). Using the conserved domain sequence of *clpP* as a query, we did not find any evidence of divergent *clpP*-like ORFs that were intact. Only two lineages contained the *clpP* gene, showing dynamics in intron content. The phylogenetic distribution showed the evolutionary history of intron loss events in the subfamily Valerianoideae, in which the loss of the second intron in the common ancestor of *Fedia* and *Valerianella*, followed by loss of the first intron independently in *Valerianella* (Figure 5). The subfamily Caprifolioideae plastome contains intronless *clpP*, but the history of loss events is unclear from the present data. Direct genomic deletion, exonization of introns, retroprocessing, and gene conversion with foreign copies are possible mechanisms of intron loss [44]. Intensified positive or negative selection in the *clpP* gene among Caprifoliaceae *s.l.* suggests that positive selection may act on the nuclear-encoded plastid-targeted genes. Cytonuclear coevolution between plastid- and nuclear-encoded subunits in Caprifoliaceae *s.l.* should be explored to test this hypothesis.

4. Materials and Methods

4.1. Genome Sequencing, Assembly, and Annotation

Total genomic DNA (gDNA) from *V. fauriei* and *V. locusta* from the subfamily Vaerianoideae, as well as *D. japonicus*, and *S. comosa* from the subfamily Dipsacoideae (Table S5), were isolated from fresh leaf tissues of a single individual using the Exgene Plant SV Mini Kit (GeneAll, Seoul, South Korea) following the manufacturer's protocol. The gDNA of *F. cornucopiae* was provided by the Royal Botanic Gardens Kew DNA and Tissue Collection (Table S5). The gDNAs were sequenced using an Illumina Hiseq2500 sequencing platform (Illumina, San Diego, CA, USA), generating 6 Gb of 150 bp paired-end (PE) reads from a 550 bp insert library.

The PE reads were assembled de novo using Velvet v1.2.10 [45] using multiple k-mers (99 to 141). For each plastome, the longest contigs that reflected a complete plastome with only one copy of the IR were aligned manually, and the consensus was taken as the final genome sequence. Finished plastomes were annotated using a BLAST-like algorithm in Geneious Prime 2021.1.1 (www.geneious.com, accessed on 18 August 2021) with the genes of Nicotiana tabacum L. plastome (NC_001879) as the reference, and the open reading frames (ORFs) were confirmed using the "Find ORFs" option. Circular plastome maps were drawn using OrganellarGenomeDRAW (OGDRAW) v1.3.1 (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html, accessed on 18 August 2021) [46]. The plastomes were deposited in GenBank (accession numbers MZ934745-MZ934749).

4.2. Comparative Analyses

Repetitive DNA sequences in each plastome were identified by performing "blastn" searches using BLAST + v2.6.0 [47] against itself, with a word size of 11, an e-value of 1×10^{-6}. The newly sequenced plastomes and the 19 published Caprifoliaceae *s.l.* plastomes were aligned with the outgroup *Viburnum betulifolium* Batalin from Adoxaceae using the "progressiveMauve" algorithm in Mauve v2.3.1 [48] in Geneious Prime. The National

Center for Biotechnology Information (NCBI) Conserved Domain Database (CDD) v3.19 was used for functional domain annotation (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi, accessed on 18 August 2021) [49]. Transcriptome from *Dipsacus asperoides* C.Y.Cheng & T.M.Ai was assembled de novo with Trinity [50] using the Sequence Read Archive (SRA) (SRR2043985). The potential nuclear-encoded transcript was identified in the transcriptome by using "blastn" (e-value cutoff of 1×10^{-10}) with the plastid-encoded *rps15* gene sequence from *Triplostegia glandulifera* Wall. ex DC. as a query. Chloroplast transit peptide (cTP) was predicted by TargetP v1.1 [51].

4.3. Estimation of Substitution Rates

In total, 72 plastid protein-coding genes shared by all selected 29 taxa from newly sequenced plastomes and from the published plastomes were sampled (Table S1). Individual genes were aligned using the back-translation method with MAFFT [52] in Geneious Prime. The constraint tree was generated using the maximum likelihood method in IQ-TREE v2.1.2 [53] with concatenated sequence alignment. To estimate the rates of nucleotide substitution, all genes and the concatenated sequences were analyzed individually. The nonsynonymous (d_N) and synonymous (d_S) substitution rates for datasets were calculated in PAML v4.8 [54] with the constraint tree. Codon frequencies were estimated using the F3 $\times$ 4 model. LRTs were performed in Hyphy v2.5.23(MP) [55] to test d_N/d_S changes using the MG94xREV codon model. To test for potential relaxed selection, the RELAX [56] implemented in HyPhy was used on the Datamonkey Adaptive Evolution Server (https://www.datamonkey.org/, accessed on 18 August 2021) [57].

4.4. Survey of Variability in the Plastid-Encoded accD Gene

To examine length variation in the *accD* gene at the inter- and intraspecific levels, 38 *V. fauriei* individuals and 12 *V. sambucifolia* f. *dageletiana* individuals were sampled (Table S5). The gDNAs were extracted from the fresh leaves or herbarium specimens using the GeneAll Kit, or the methods described by Allen et al. [58]. Variable regions in the *accD* gene were amplified by PCR using specific primers designed with Primer3 in Geneious Prime (64F: 5′-AACTCTTATGATTCGGTTTCTCGT-3′ and 1328R: 5′-ATACCGGTTTGAATAGCCTCAGTT-3′). Each reaction was 50 μL in volume, including 38.75 μL of distilled water, 5 μL of 10 × Taq Reaction Buffer, 1 μL of dNTPs (10 mM), 0.25 μL of DiaStar^TM Taq polymerase (5 units/μL, Solgent Co., Daejeon, Korea), 1 μL of each primer (10 pmole/μL), and 1 μL of total gDNA (20 ng). All reactions consisted of included an initial denaturation step (95 °C for 2 min), 35 cycles of denaturation (95 °C for 20 s), annealing (60 °C for 40 s), and extension (72 °C for 1 min 30 s), followed by a final extension (72 °C for 5 min). The PCR products were purified using a PCR purification kit (MGmed, Korea) according to the manufacturer's protocol. Sequencing of PCR products was carried out using an ABI 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA, USA) at Solgent Co. The nucleotide sequences of the plastid *accD* copies were aligned using MUSCLE [59] in Geneious Prime.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ijms221910485/s1, Figure S1: Maps for the newly sequenced plastomes, Figure S2: Nucleotide and amino acid sequences of the nuclear-encoded *RPS15* gene from *Dipsacus*, Figure S3: Duplication of the *trnE-UUC* gene in *Dipsacus japonicus* plastome, Figure S4: Structural alignments of Caprifoliaceae *s.l.* plastomes using Mauve. Figure S5: Amino acid sequence alignments of the plastid-encoded *accD* of Caprifoliaceae *s.l.*, Figure S6. Boxplots of the values of nonsynonymous and synonymous substitution rates of the plastid-encoded *accD* for Caprifoliaceae *s.l.* and outgroups, Table S1: GenBank accession numbers for taxa used in this study, Table S2: Pairwise Wilcoxon rank-sum tests of d_N and d_S values among plastid genes within Caprifoliaceae *s.l.*, Table S3: Positive selection on Caprifoliaceae *s.l.* plastid genes, Table S4: CD-search results of the plastid-encoded *accD* gene from Caprifoliaceae *s.l.*, Table S5: Material information and GenBank accession numbers for length variation in the plastid-encoded *accD* gene.

References

1. Ruhlman, T.A.; Jansen, R.K. The plastid genomes of flowering plants. *Methods Mol. Biol.* **2014**, *1132*, 3–38. [CrossRef] [PubMed]
2. Jansen, R.K.; Cai, Z.; Raubeson, L.A.; Daniell, H.; dePamphilis, C.W.; Leebens-Mack, J.; Müller, K.F.; Guisinger-Bellian, M.; Haberle, R.C.; Hansen, A.K.; et al. 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*, 19369–19374. [CrossRef] [PubMed]
3. Timmis, J.N.; Ayliffe, M.A.; Huang, C.Y.; Martin, W. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **2004**, *5*, 123–135. [CrossRef] [PubMed]
4. Cosner, M.E.; Raubeson, L.A.; Jansen, R.K. Chloroplast DNA rearrangements in Campanulaceae: Phylogenetic utility of highly rearranged genomes. *BMC Evol. Biol.* **2004**, *4*, 1–17. [CrossRef] [PubMed]
5. Sloan, D.B.; Triant, D.A.; Forrester, N.J.; Bergner, L.M.; Wu, M.; Taylor, D.R. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe Sileneae (Caryophyllaceae). *Mol. Phylogenetics Evol.* **2014**, *72*, 82–89. [CrossRef] [PubMed]
6. Cai, Z.; Guisinger, M.; Kim, H.-G.; Ruck, E.; Blazier, J.C.; McMurtry, V.; Kuehl, J.V.; Boore, J.; Jansen, R.K. Extensive reorganization of the plastid genome of *Trifolium subterraneum* (Fabaceae) is associated with numerous repeated sequences and novel DNA insertions. *J. Mol. Evol.* **2008**, *67*, 696–704. [CrossRef] [PubMed]
7. Guisinger, M.M.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: Rearrangements, repeats, and codon usage. *Mol. Biol. Evol.* **2011**, *28*, 583–600. [CrossRef] [PubMed]
8. Lee, H.-L.; Jansen, R.K.; Chumley, T.W.; Kim, K.-J. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol. Biol. Evol.* **2007**, *24*, 1161–1180. [CrossRef] [PubMed]
9. Park, S.; An, B.; Park, S. Reconfiguration of the plastid genome in *Lamprocapnos spectabilis*: IR boundary shifting, inversion, and intraspecific variation. *Sci. Rep.* **2018**, *8*, 13568. [CrossRef]
10. Xu, X.; Wang, D. Comparative Chloroplast Genomics of *Corydalis* Species (Papaveraceae): Evolutionary Perspectives on Their Unusual Large Scale Rearrangements. *Front. Plant Sci.* **2021**, *11*. [CrossRef]
11. Guisinger, M.M.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 18424–18429. [CrossRef] [PubMed]
12. Schwarz, E.N.; Ruhlman, T.A.; Weng, M.-L.; Khiyami, M.A.; Sabir, J.S.M.; Hajarah, N.H.; Alharbi, N.S.; Rabah, S.O.; Jansen, R.K. Plastome-Wide Nucleotide Substitution Rates Reveal Accelerated Rates in Papilionoideae and Correlations with Genome Features Across Legume Subfamilies. *J. Mol. Evol.* **2017**, *84*, 187–203. [CrossRef] [PubMed]
13. Shrestha, B.; Weng, M.-L.; Theriot, E.C.; Gilbert, L.E.; Ruhlman, T.A.; Krosnick, S.E.; Jansen, R.K. Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba. Mol. Phylogenetics Evol.* **2019**, *138*, 53–64. [CrossRef]
14. Weng, M.-L.; Blazier, J.C.; Govindu, M.; Jansen, R.K. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol. Biol. Evol.* **2014**, *31*, 645–659. [CrossRef]
15. Park, S.; Ruhlman, T.A.; Weng, M.-L.; Hajrah, N.H.; Sabir, J.S.M.; Jansen, R.K. Contrasting Patterns of Nucleotide Substitution Rates Provide Insight into Dynamic Evolution of Plastid and Mitochondrial Genomes of *Geranium. Genome Biol. Evol.* **2017**, *9*, 1766–1780. [CrossRef]
16. Magee, A.M.; Aspinall, S.; Rice, D.W.; Cusack, B.P.; Sémon, M.; Perry, A.S.; Stefanović, S.; Milbourne, D.; Barth, S.; Palmer, J.D.; et al. Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res.* **2010**, *20*, 1700–1710. [CrossRef]

17. Blazier, J.C.; Ruhlman, T.A.; Weng, M.-L.; Rehman, S.K.; Sabir, J.S.M.; Jansen, R.K. Divergence of RNA polymerase α subunits in angiosperm plastid genomes is mediated by genomic rearrangement. *Sci. Rep.* **2016**, *6*, 24595. [CrossRef]

18. Barnard-Kubow, K.B.; Sloan, D.B.; Galloway, L.F. Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evol. Biol.* **2014**, *14*, 268. [CrossRef] [PubMed]

19. Parkinson, C.L.; Mower, J.P.; Qiu, Y.-L.; Shirk, A.J.; Song, K.; Young, N.D.; Claude, W.d.; Palmer, J.D. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. *BMC Evol. Biol.* **2005**, *5*, 1–12. [CrossRef]

20. Rousseau-Gueutin, M.; Huang, X.; Higginson, E.; Ayliffe, M.; Day, A.; Timmis, J.N. Potential Functional Replacement of the Plastidic Acetyl-CoA Carboxylase Subunit (*accD*) Gene by Recent Transfers to the Nucleus in Some Angiosperm Lineages. *Plant Physiol.* **2013**, *161*, 1918–1929. [CrossRef] [PubMed]

21. Molina, J.; Hazzouri, K.M.; Nickrent, D.; Geisler, M.; Meyer, R.S.; Pentony, M.M.; Flowers, J.M.; Pelser, P.; Barcelona, J.; Inovejas, S.A.; et al. Possible Loss of the Chloroplast Genome in the Parasitic Flowering Plant *Rafflesia lagascae* (Rafflesiaceae). *Mol. Biol. Evol.* **2014**, *31*, 793–803. [CrossRef] [PubMed]

22. Sabir, J.; Schwarz, E.; Ellison, N.; Zhang, J.; Baeshen, N.A.; Mutwakil, M.; Jansen, R.; Ruhlman, T. Evolutionary and biotechnology implications of plastid genome variation in the inverted-repeat-lacking clade of legumes. *Plant Biotechnol. J.* **2014**, *12*, 743–754. [CrossRef] [PubMed]

23. Babiychuk, E.; Vandepoele, K.; Wissing, J.; Garcia-Diaz, M.; De Rycke, R.; Akbari, H.; Joubès, J.; Beeckman, T.; Jänsch, L.; Frentzen, M.; et al. Plastid gene expression and plant development require a plastidic protein of the mitochondrial transcription termination factor family. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6674–6679. [CrossRef]

24. Schulte, W.; Töpfer, R.; Stracke, R.; Schell, J.; Martini, N. Multi-functional acetyl-CoA carboxylase from *Brassica napus* is encoded by a multi-gene family: Indication for plastidic localization of at least one isoform. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 3465–3470. [CrossRef]

25. Konishi, T.; Shinohara, K.; Yamada, K.; Sasaki, Y. Acetyl-CoA Carboxylase in Higher Plants: Most Plants Other Than Gramineae Have Both the Prokaryotic and the Eukaryotic Forms of This Enzyme. *Plant Cell Physiol.* **1996**, *37*, 117–122. [CrossRef] [PubMed]

26. Gornicki, P.; Faris, J.; King, I.; Podkowinski, J.; Gill, B.; Haselkorn, R. Plastid-localized acetyl-CoA carboxylase of bread wheat is encoded by a single gene on each of the three ancestral chromosome sets. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 14179–14184. [CrossRef]

27. Rockenbach, K.; Havird, J.C.; Monroe, J.G.; Triant, D.A.; Taylor, D.R.; Sloan, D.B. Positive Selection in Rapidly Evolving Plastid–Nuclear Enzyme Complexes. *Genetics* **2016**, *204*, 1507–1522. [CrossRef] [PubMed]

28. Williams, A.M.; Friso, G.; van Wijk, K.J.; Sloan, D.B. Extreme variation in rates of evolution in the plastid Clp protease complex. *Plant J.* **2019**, *98*, 243–259. [CrossRef]

29. White, S.W.; Zheng, J.; Zhang, Y.M.; Rock, C.O. The structural biology of type II fatty acid biosynthesis. *Annu. Rev. Biochem.* **2005**, *74*, 791–831. [CrossRef]

30. Wakil, S.J.; Stoops, J.K.; Joshi, V.C. Fatty acid synthesis and its regulation. *Annu. Rev. Biochem.* **1983**, *52*, 537–579. [CrossRef]

31. Adam, Z.; Adamska, I.; Nakabayashi, K.; Ostersetzer, O.; Haussuhl, K.; Manuell, A.; Zheng, B.; Vallon, O.; Rodermel, S.R.; Shinozaki, K.; et al. Chloroplast and Mitochondrial Proteases in *Arabidopsis*. A Proposed Nomenclature1. *Plant Physiol.* **2001**, *125*, 1912–1918. [CrossRef] [PubMed]

32. Shikanai, T.; Shimizu, K.; Ueda, K.; Nishimura, Y.; Kuroiwa, T.; Hashimoto, T. The Chloroplast *clpP* Gene, Encoding a Proteolytic Subunit of ATP-Dependent Protease, is Indispensable for Chloroplast Development in Tobacco. *Plant Cell Physiol.* **2001**, *42*, 264–273. [CrossRef] [PubMed]

33. Christenhusz, M.J.; Byng, J.W. The number of known plants species in the world and its annual increase. *Phytotaxa* **2016**, *261*, 201–217. [CrossRef]

34. Wang, H.-X.; Liu, H.; Moore, M.J.; Landrein, S.; Liu, B.; Zhu, Z.-X.; Wang, H.-F. Plastid phylogenomic insights into the evolution of the Caprifoliaceae *s.l.* (Dipsacales). *Mol. Phylogenetics Evol.* **2020**, *142*, 106641. [CrossRef] [PubMed]

35. Wang, H.-X.; Morales-Briones, D.F.; Moore, M.J.; Wen, J.; Wang, H.-F. A phylogenomic perspective on gene tree conflict and character evolution in Caprifoliaceae using target enrichment data, with Zabelioideae recognized as a new subfamily. *J. Syst. Evol.* **2021**, *59*, 897–914. [CrossRef]

36. Fan, W.-B.; Wu, Y.; Yang, J.; Shahzad, K.; Li, Z.-H. Comparative Chloroplast Genomics of Dipsacales Species: Insights Into Sequence Variation, Adaptive Evolution, and Phylogenetic Relationships. *Front. Plant Sci.* **2018**, *9*. [CrossRef]

37. Goulding, S.E.; Wolfe, K.; Olmstead, R.; Morden, C. Ebb and flow of the chloroplast inverted repeat. *Mol. Gen. Genet. MGG* **1996**, *252*, 195–206. [CrossRef]

38. Haberle, R.C.; Fourcade, H.M.; Boore, J.L.; Jansen, R.K. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J. Mol. Evol.* **2008**, *66*, 350–361. [CrossRef]

39. Gemayel, R.; Vinces, M.D.; Legendre, M.; Verstrepen, K.J. Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. *Annu. Rev. Genet.* **2010**, *44*, 445–477. [CrossRef]

40. 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*, 417–427. [CrossRef]

41. Erixon, P.; Oxelman, B. Whole-Gene Positive Selection, Elevated Synonymous Substitution Rates, Duplication, and Indel Evolution of the Chloroplast *clpP1* Gene. *PLoS ONE* **2008**, *3*, e1386. [CrossRef]

42. Wang, W.-C.; Chen, S.-Y.; Zhang, X.-Z. Chloroplast Genome Evolution in Actinidiaceae: *clpP* Loss, Heterogenous Divergence and Phylogenomic Practice. *PLoS ONE* **2016**, *11*, e0162324. [CrossRef]

43. Yao, X.; Tang, P.; Li, Z.; Li, D.; Liu, Y.; Huang, H. The First Complete Chloroplast Genome Sequences in Actinidiaceae: Genome Structure and Comparative Analysis. *PLoS ONE* **2015**, *10*, e0129347. [CrossRef] [PubMed]

44. Hepburn, N.J.; Schmidt, D.W.; Mower, J.P. Loss of two introns from the *Magnolia tripetala* mitochondrial *cox2* gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss. *Mol. Biol. Evol.* **2012**, *29*, 3111–3120. [CrossRef]

45. Zerbino, D.R.; Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **2008**, *18*, 821–829. [CrossRef]

46. 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–W64. [CrossRef] [PubMed]

47. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [CrossRef] [PubMed]

48. Darling, A.C.E.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [CrossRef] [PubMed]

49. Marchler-Bauer, A.; Lu, S.; Anderson, J.B.; Chitsaz, F.; Derbyshire, M.K.; DeWeese-Scott, C.; Fong, J.H.; Geer, L.Y.; Geer, R.C.; Gonzales, N.R.; et al. CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **2010**, *39*, D225–D229. [CrossRef] [PubMed]

50. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef] [PubMed]

51. Emanuelsson, O.; Brunak, S.; von Heijne, G.; Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2007**, *2*, 953–971. [CrossRef] [PubMed]

52. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]

53. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef] [PubMed]

54. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [CrossRef]

55. Kosakovsky Pond, S.L.; Poon, A.F.Y.; Velazquez, R.; Weaver, S.; Hepler, N.L.; Murrell, B.; Shank, S.D.; Magalis, B.R.; Bouvier, D.; Nekrutenko, A.; et al. HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Mol. Biol. Evol.* **2019**, *37*, 295–299. [CrossRef]

56. Wertheim, J.O.; Murrell, B.; Smith, M.D.; Kosakovsky Pond, S.L.; Scheffler, K. RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. *Mol. Biol. Evol.* **2014**, *32*, 820–832. [CrossRef]

57. Weaver, S.; Shank, S.D.; Spielman, S.J.; Li, M.; Muse, S.V.; Kosakovsky Pond, S.L. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol. Biol. Evol.* **2018**, *35*, 773–777. [CrossRef] [PubMed]

58. Allen, G.; Flores-Vergara, M.; Krasynanski, S.; Kumar, S.; Thompson, W. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* **2006**, *1*, 2320–2325. [CrossRef]

59. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]