Phylogeny and adaptive evolution of subgenus Rhizirideum (Amaryllidaceae, Allium) based on plastid genomes

Xiao Fu
College of Life Sciences, Sichuan University
Deng-Feng Xie
College of Life Sciences, Sichuan University
Yu-Yang Zhou
College of Life Sciences, Sichuan University
Rui-Yu Cheng
College of Life Sciences, Sichuan University
Xiang-Yi Zhang
College of Life Sciences, Sichuan University
Song-dong Zhou
College of Life Sciences, Sichuan University
Xing-Jin He (xjhe@scu.edu.cn)
College of Life Sciences, Sichuan University

Research Article

Keywords: Allium, Subgen. Rhizirideum, Plastid genomes, Adaptive evolution, Phylogeny

Posted Date: August 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1827197/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on February 1st, 2023. See the published version at https://doi.org/10.1186/s12870-022-03993-z.

Abstract

The subgenus *Rhizirideum* in genus *Allium* consists of 38 species across the world. Previous studies on this subgenus mainly focused on separate sections. To investigate the adaptive evolution and phylogenetic relationships of this subgenus, we selected eleven species of subgen. *Rhiziridem* to conduct the comparative plastome analysis. As a result, the *Rhizirideum* plastomes were relatively conservative in terms of structure, IR/SC borders, codon usage, and repeat sequence. Unlike the previous reports, *A. polyrhizum*, representing section *Caespitosoprason*, was resolved as the basal taxon of subgen. *Rhizirideum*. Most of the single-copy CDSs were under purifying selection, while eleven genes presented relaxed selection, such as *atpB*, *psbD* and *rbcL*, which were mainly involved in the photosynthesis and respiration processes. The pseudogenization and relaxed selection of genes, the introns, the slight shift of IR/SC borders and the low GC content can be ragarded as the adaptation of plants to the environment. Our study highlighted the advantage of plastid genomes in reconstructing phylogenetic relationships. Also, we considered that section *Caespitosoprason* needed to be separated from section *Rhizomatosa*. In summary, our research provided new insights into the phylogeny, adaptive evolution, and taxonomy of subgen. *Rhizirideum*.

1. Introduction

Allium (Allioideae, Amaryllidaceae), one of the largest genera of monocots, has more than 900 verified species on the Earth (Herden et al., 2016). Many species in this genus have been used for edible (e.g., *A. sativum*, *A. tuberosum*, *A. porrum*), medicinal (e.g., *A. sativum*, *A. victorialis*, *A. cepa*) and ornamental (e.g., *A. giganteum*, *A. wallichii*, *A. moly*). Actually, genus *Allium* was originally established by Linnaeus in *Species plantarum* (Linnaeus, 1753), which initially contained only 30 *Allium* species sorted into 3 alliances. Subsequently, many scientists published a large quantity of new *Allium* taxa, and analyses on the taxonomy and phylogeny of *Allium* also emerged because of the complicated relationship within this genus. Then, Regel's monograph included 263 species and distributed them into 6 sections (Regel, 1875, 1887). Afterwards, Traub (1968) sorted 600 *Allium* species into 3 subgenera, including 36 sections and subsections. And Wendelbo (1969) first proposed the subgenus *Rhizirideum*. After that, Kamelin (1973) revised the phylogeny of *Allium* and classified it into 6 subgenera (44 sections & subsections). In Kamelin's taxonomy, the subgenus *Rhizirideum* contained 150 species, such as *A. cepa*, *A. senescens*, and *A. ramosum* and further sorted them into 12 sections and subsections as Sect. *cepa*, Sect. *Butomissa*, Sect. *Rhizirideum*. Later, Friesen et al. (2006) reconstructed the phylogeny of *Allium* based on ITS data and divided it into three main evolutionary lineages. Friesen et al. (2006) put the new subgenus *Rhizirideum* forward (Subgen. *Rhizirideum* in the following) and distributed approximately 780 *Allium* species into 15 genera (72 sections). At the same time, the previous subgenus *Rhizirideum* was disproved and found to be nonmonophyletic. Sixteen species (e.g., A. *senescens*) in the previous subgenus *Rhizirideum* were still assorted into the new one while the others were assorted into other subgenera such as *Anguinum* (e.g., A. *victorialis*), *Cepa* (e.g., A. *cepa*), and *Butomiss*a (e.g., A. *ramosum*). And subgen. *Rhizirideum* included five sections: *Rhizirideum*, *Rhizomatosa*, *Tenuissima*, *Eduardii*, and *Caespitosoprason*. Recently, Li et al. (2016) provided adequate evidence for the monophyly of subgen. *Rhizirideum* based on chloroplast DNA fragments data. Friesen et al. (2020) then merged section *Caespitosoprason* into section *Rhizomatosa* under subgen. *Rhizirideum.*

At present, subgen. *Rhizirideum* consists of four sections (*Rhizirideum*, *Rhizomatosa*, *Tenuissima*, *Eduardii*) and thirty-eight species in total (Friesen et al., 2006; Jang et al., 2021). It was located in the third lineage of the *Allium* phylogeny. Species in this subgenus were characterized by obvious rhizome, leaves subcylindrical to flat, perianth white to purple, ovary with 2 ovules per locule, and inner filaments broadened at the base (Fig. 1, Figure S1).

Studies of species in subgen. *Rhizirideum* have been conducted frequently in the past century (Shopova, 1966; Friesen, 1992; Dubouzet et al., 1997; Do and Seo, 2000; Raamsdonk et al., 2000; Kim et al., 2001; Lee, 2001; Raamsdonk et al., 2003; Ricroch et al., 2005; Friesen et al., 2006; Özler and Pehlivan, 2010; Rola, 2014; Li et al., 2016; Friesen et al., 2020). It was found that the chromosome base number of this subgenus was eight and the ploidy was mainly 2x or 4x. And Species in this subgenus spread over the Eurasian steppe. Sinitsyna et al. (2016) divided section *Rhizirideum* into two geographical groups, the Asiatic and European groups. The diversification and speciation of this section coincide with the history of the modern Eurasian

steppe. Meanwhile, the latest study of section *Rhizomatosa* on biogeography indicated that species in this section were distributed in the Central Asian steppe and the distribution was in accordance with the history of the landscape and climate (Friesen et al., 2020). However, some of the phylogenetic studies were focused on the previous subgenus *Rhizirideum* and the others were focused on section *Rhizirideum* and section *Rhizomatosa*. Phylogenetic analysis on section *Tenuissima* and section *Eduardii* were lacking so that more fieldwork and further analysis should be undertaken.

In recent years, the complete chloroplast genome has been popular for its conservative structure, low recombination rate, and enormous genetic information. It has been widely used in the field of phylogenetic reconstruction and adaptive evolution (Li et al., 2021; Tang et al., 2021; Wen et al., 2021). Several *Allium* taxa have also been studied on their plastomes, for instance, section C*epa*, *Daghestanica*, and subgenus *Cyathophora* (Xie et al., 2019; Yang et al., 2020; Zyab et al., 2021). Xie et al. (2020) reconstructed the phylogenetic relationship of the genus *Allium* with thirty-nine complete chloroplast genomes and revealed the evolutionary features of *Allium*. However, similar studies on the subgen. *Rhizirideum* have not yet been reported. Previous phylogenetic studies of subgen. *Rhizirideum* were mostly based on ITS or plastid DNA fragments, which provided limited information for infrageneric relationships. Furthermore, analysis of adaptive evolution was also inadequate. Thus it is necessary to further investigate the composition, structure, and evolution of subgen. *Rhizirideum* plastomes. We collected eleven species in Subgen. *Rhizirideum* and combined twenty-two related species to conduct comparative chloroplast genome analyses. Our aims are as follows: (1) to compare the structures and genetic compositions of plastomes of 11 *Rhizirideum* species and (2) to reconstruct the phylogeny of subgen. *Rhizirideum*; (3) to analyze the adaptive evolution of subgen. *Rhizirideum* species.

2. Results

2.1 Plastome structure of subgen. Rhizirideum species

The subgen. *Rhizirideum* plastomes shared a quadripartite circular structure with two inverted repeats (IRa & IRb), one large single copy (LSC), and one small single copy (SSC) (Fig. 2, Table 1). The sizes of eleven *Rhizirideum* plastomes ranged from 153,667 bp to 153,257 bp, and the overall GC content of them ranged from 36.8–36.9%. Each plastome contained 141 genes, among which there were 87 or 89 protein-coding sequences (9 or 10 in IRs), 30 tRNA-coding genes (8 in IRs), and 4 rRNA-coding genes (4 in IRs). Moreover, 26 genes were found to be interpreted by introns (Table 2). The gene *clpP*, *rps12*, and *ycf3* had two introns inserted into their sequences. And the *trnK-UUU* gene had the longest intron, where the *matK* gene was located. The *rps12* is a trans-spliced gene with the 5'-end in the LSC region and the duplicated 3'-ends in the IR regions.

Table 1
Summary of the subgenus *Rhizirideum* plastomes

Taxon	Total genome length(bp)	GC content(%)	IR length (bp)	LSC length (bp)	SSC length (bp)	Gene Number	Protein-coding genes	tRNAs	rRNAs
Allium spirale	153,549	36.8	26,493	82,576	17,987	141	89(10)	38(8)	8(4)
Allium senescens	153,516	36.8	26,491	82,548	17,986	141	89(10)	38(8)	8(4)
Allium nutans	153,456	36.9	26,487	82,531	17,951	141	87(9)	38(8)	8(4)
Allium mongolicum	153,667	36.8	26,490	82,645	18,042	141	87(9)	38(8)	8(4)
Allium polyrhizum	153,614	36.8	26,450	82,624	18,090	141	87(9)	38(8)	8(4)
Allium bidentatum	153,443	36.8	26,459	82,504	18,021	141	89(10)	38(8)	8(4)
Allium caespitosum	153,667	36.8	26,490	82,643	18,044	141	87(9)	38(8)	8(4)
Allium tenuissimum	153,459	36.8	26,491	82,484	17,993	141	87(9)	38(8)	8(4)
Allium anisopodium	153,407	36.8	26,491	82,426	17,999	141	87(9)	38(8)	8(4)
Allium eduardii	153,497	36.9	26,732	82,296	17,737	141	89(10)	38(8)	8(4)
Allium przewalskianum	153,257	36.9	26,437	82,410	17,973	141	89(10)	38(8)	8(4)

*Numbers in brackets note the number of genes in IRs.

Table 2
Summary of genes interrupted by introns in *Rhizirideum* plastomes

No.	Gene	Region	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
1	*atpF*	LSC	144+	789	411+		
2	*clpP*	LSC	69+	1094	294+	879	252+
3	*ndhA*	SSC	558+	1128	540+		
4	*ndhB*	IRb	777+	701	756+		
5	*ndhB*	IRa	777−	701	756−		
6	*petB*	LSC	6−	920	642−		
7	*petD*	LSC	8−	746	514−		
8	*rpl16*	LSC	9+	1042	399+		
9	*rpl2*	IRb	387+	659	432+		
10	*rpl2*	IRa	387−	659	432−		
11	*rpoC1*	LSC	432+	759	1623+		
12	*rps12a*	LSC, IRa	114+	69984	232−	542	26−
13	*rps12b*	LSC,IRb	114+	28994	232+	542	26+
14	*rps16*	LSC	40+	846	197+		
15	*trnA-UGC*	IRb	38−	815	35+		
16	*trnA-UGC*	IRa	38+	815	35+		
17	*trnG-TCC*	LSC	23−	692	49−		
18	*trnI-GAU*	IRb	37−	934	35−		
19	*trnI-GAU*	IRa	37+	934	35+		
20	*trnL-UAA*	LSC	35−	305	50−		
21	*trnV-UAC*	LSC	37+	598	37+		
22	*ycf3*	LSC	129+	722	228+	738	
23	*ycf68*	IRb	42−	31	411−		153+
24	*ycf68*	IRa	42+	31	411+		
25	*trnK-UUU*	LSC	37+	2564	35+		

*The data in this table is from *Allium senescens* plastome.

Multiple alignments of *Rhizirideum* plastomes showed similar structural features. Equivalent distribution patterns of GC islands were displayed among eleven *Rhizirideum* plastomes (Fig. 2, rings a-b). LSC and SSC regions, especially the former, showed lower sequence identities than those of IR regions. In the single copy (SC) regions, *Rhizirideum* species share several

divergent sequence sites. For instance, *A. anisopodium* and *A. tenuissimum* from Sect. *Tenuissima* (Fig. 2, rings g-h), as well as *A. eduardii* and *A. przewalskianum* from Sect. *Eduardia* (Fig. 2, rings l-m) shared five highly divergent sites with *A. senescens* in the LSC region (*psbK ~ trnS-GCU*, *trnT-GGU ~ psbD*, *trnP-UGG ~ psaJ*, *clpP ~ psbB*, *rps8 ~ rpl14*) and one in the SSC region (*rpl32 ~ trnL-UAG*). Besides, diagram drawn by mVISTA (Fig. 3) showed sequence identities of different regions in *Rhizirideum* plastomes straightforward regarding *A. senescens* as a reference. As indicated, exon regions had higher identity values than UTR and CNS regions. IR regions also had higher sequence identities than SC regions.

We selected 112 genes (Fig. 4A) and 100 intergenetic regions (Fig. 4B) to compute their nucleotide diversity (Pi) values by using DNasp software. As the results indicated, the average Pi value of the genes (0.0191) was smaller than that of the intergenetic regions (0.0421). In terms of Pi values, the top three genes were *rpl32* (0.0859), *rps16* (0.0406), and *ndhF* (0.03934), while the top three intergenetic regions were *rps15-ycf1* (0.0775), *ndhF-rpl32* (0.0757) and *ccsA-ndhD* (0.0717).

2.2 IR/SC borders

Lengths of the IR and SC regions of the eleven subgen. *Rhizirideum* plastomes were compared (Fig. 5). In the results, *A.eduardii* had the longest IR sequence (26,732 bp) while the last three IRs in terms of size belonged to *A. przewalskianum* (26,437 bp), *A. polyrhizum* (26,450 bp), and *A. bidentatum* (26,459 bp). For SSC regions, *A. polyrhizum* had the largest one (18,090 bp) while *A. eduardii* (17,737 bp) and *A. przewalskianum* (17,973 bp) got the two smallest ones. The longest LSC region belonged to *A. mongolicum* and the shortest three belonged to *A. eduardii* (82,296 bp), *A. anisopodium* (82,426 bp), and *A. przewalskianum* (82,410 bp).

Then, the position of the IR/SC borders was examined (Fig. 5). The gene content on both sides of the IR/SC borders of eleven *Rhizirideum* plastomes were conserved by and large. The LSC/IRb border was *rps19*/*rpl22*. Mostly, *rpl22* was interrupted by LSC/IRb borders, and *rps19* was no less than 63 bp away from IRa/LSC borders. There were exceptions anyway. The *rpl22* gene of *A. bidentatum* plastome was completely located in the LSC region (18 bp from the LSC/IRb border) and the *rps19* gene of *A. polyrhizum* was just 47 bp away from the IRb/LSC border. In addition to IRb/LSC borders, *rps19*/*psbA* genes were 30bp/146bp away from the IRa/LSC border in the *A. bidentatum* plastome, varying from those in another ten *Rhizirideum* plastomes (63 ~ 82 bp and 67 ~ 85 bp, respectively).

For SSC bounderies, two SSC/IR borders crossed two *ycf1* genes in most of the *Rhizirideum* plastomes. On the subject of IRb/SSC borders, a large part of *ycf1b* sequence was mostly located in the IRb region, while gene *ndhF* was completely located in the SSC region. Nevertheless, there were still several exceptions. The whole *ycf1b* gene of the *A. eduardii* plastome was located in the IRb region (away from the IRb/SSC border by 208 bp). In plastomes of *A. anisopodium* and *A. tenuissimum*, IRb/SSC borders overlapped *ndhF* genes by 7 bp and 8 bp, respectively. Gene *ycf1*a of *Rhizirideum* was 5,295 bp in length, except for *A. anisopodium* and *A. tenuissimum* (5,313 bp). Concerning SSC/IRa borders, gene *ycf1*a was divided into two fragments, and its IRa side ranged from 679 bp (*A. eduardii*) to 1,309 bp (*A. przewalskianum*).

2.3 Codon usage bias analysis

The codon usage bias of eleven subgen. *Rhizirideum* plastomes was analyzed by using program condonW (Fig. 6). The total number of codons in the protein-coding regions of each plastome ranged from 16,566 (*A. przewalskianum*) to 16,617 (*A. spirale*), with relative synonymous codon usage (RSCU) ranging from 2.29 (UUA) to 0.27 (CUC). In terms of the codons of eleven plastomes, six synonymous codons coding leucine (Leu) held the largest proportion (10.5%), and two for cysteine (Cys) held the smallest proportion (1.09%) except for stop codons (0.41%). Methionine (Met, AUG) and tryptophan (Trp, UGG) showed no codon bias and were encoded by only one codon. Thirty-one frequently used codons with RSCU > 1 encoded all kinds of amino acids except for Trp and Met. Of the codons used frequently, only 25.8% were ended by G/C. The GC content for the first, second, and third codon positions averaged 47.3%, 39.3%, and 27.2%, respectively.

2.4 Repeat sequence analysis

We detected 879 simple sequence repeats (SSRs) in the eleven *Rhizirideum* plastomes (Fig. 7B). *A. mongolicum* and *A. caespitosum* contained the most SSRs (88) whereas *A. spirale* and *A. polyrhizum* (74) contained the least. SSRs with 1 bp ~ 5 bp could be witnessed across the eleven plastomes, but those with 6 bp (i.e., the hexanucleotides) were scarce. Among all kinds of SSRs, mononucleotides (55.97%) were the most abundant, followed by compound microsatellites (15.70%), dinucleotides (11.95%), tetranucleotides (11.04%), and pentanucleotides (1.99%). Most of the SSRs were composed of A/T as G/C rarely occurred. They were distributed more in LSC regions than in IR or SSC regions. In addition to SSRs, repeats were also detected that were 30 bp ~ 60 bp in length (Fig. 7, (A)). Four types of them were summed up to 448, including forward, reverse, palindromic and complementary repeats. The proportion of palindromic repeats (52.23%) was the highest while that of the complementary repeats (0.45%) was the lowest in all the repeats. *A. przewalskianum* contained the most repeats (49) and *A. anisopodium* contained the least repeats (31).

2.5 Phylogenetic analysis in subgen. Rhizirideum

Seventy-one plastome single-copy genes were extracted and concatenated to reconstruct the phylogenetic tree, which possessed a quite high bootstrap and posterior probability in the subgen. *Rhizirideum* branch (Fig. 8). *A. anisopodium* and *A. tenuissimum* formed a clade and belonged to section *Tenuissima*. *A. caespitosum*, *A. mongolicum* and *A. bidentatum* formed a clade and belonged to section *Rhizomatosa*. *A. senescens*, *A. spirale* and *A. nutans* clustered to form section *Rhizirideum* Clade, and *A. eduardii* and *A. przewalskianum* formed Clade *Eduardia*. Interestingly, *A. polyrhizum*, previously belonging to section *caespitosoprason*, was resolved as sister to Clade *Tenuissima* + *Rhizomatosa* + *Rhizirideum* + *Eduardia.*

2.6 Gene selective pressure

We calculated the Ka/Ks ratio (ω) of sixty-eight common protein-coding genes in eleven subgen. *Rhizirideum* plastomes and then estimated the selective pressure (Additional file 8: Table S9). Most of the genes were under purifying selection with $\omega < 0.5$. Despite this, eleven genes were found under relaxed selection with $0.5 < \omega < 1.0$ (Fig. 9). Unexpectedly, there was no gene significant ($P < 0.05$) after the likelihood ratio test (LRT). We examined the functions and relative biochemical pathways of the eleven genes mentioned above (Table 3), almost all of which were involved in the process of photosynthesis and respiration.

Table 3
Eleven genes under relaxed selection and the pathways involved in *Rhizirideum* plastomes

CDS	Ka/Ks ratio	Gene ID	Gene description	Pathway
atpB	0.60	844757	ATP synthase CF1 beta subunit	Photosynthesis (Light-Dependent Reactions)
psaB	0.51	844770	photosystem I P700 chlorophyll a apoprotein A2	Photosynthesis (Light-Dependent Reactions)
psaC	0.60	844750	photosystem I subunit VII	Photosynthesis (Light-Dependent Reactions)
psbB	0.60	844733	photosystem II 47 kDa protein	Photosynthesis (Light-Dependent Reactions)
psbD	0.64	844775	photosystem II protein D2	Photosynthesis (Light-Dependent Reactions)
psbK	0.53	844795	photosystem II protein K	Photosynthesis (Light-Dependent Reactions)
rbcL	0.62	844754	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	Calvin-Benson Cycle
ccsA	0.71	844769	cytochrome c biogenesis protein	cellular respiration
petB	0.78	844729	cytochrome b6	cellular respiration
rpl32	0.53	6161	ribosomal protein L32	protein synthesis
rpoA	0.56	844727	RNA polymerase alpha subunit	RNA synthesis

3. Discussion

3.1 Plastome structure analysis of subgen. Rhizirideum

Although events of evolution such as genome rearrangement, gene loss, IR expansion, and contraction, have been detected for many times, plastomes are generally highly conserved in genome size, structure, and gene content generally (Palmer et al., 1987; Tsudzuki et al., 1992; Lee et al., 2007; Wicke et al., 2011; Park et al., 2015; Zhu et al., 2016; Huang et al., 2017; Zhai et al., 2019; Ren et al., 2020). In this study, the subgen. *Rhizirideum* plastomes are of high conservation by large. The quantity of genes, CDSs, rRNA-coding genes, and tRNA-coding genes is 141, 87 (or 89), 8, and 38, respectively, which follows most angiosperms (Jin et al., 2014; Park et al., 2015; Huang et al., 2017; Liu et al., 2020; Ren et al., 2020; Yang et al., 2020).

At least 4 of 141 genes were pseudogenized (*orf*56, *ycf*15, *rps*2, *inf*A), while ycf68 was pseudogenized only in some species, such as *A. polyrhizum, A. nutans*, and *A. mongolicum* (Table 4). Plastome genes *ycf15*, *ycf68*, and *infA* are also pseudogenized in many other species such as *Malus pumila*, *Morus alba*, *Cynodon dactylon* (Ravi et al., 2006; Jin et al., 2014; McKain et al., 2016; Huang et al., 2017). The *rps2* gene, coding ribosomal protein S2, is lost in *Daghestanica* (Allium) plastomes but is pseudogenized in *Chlorophytum rhizopendulum* (McKain et al., 2016; Xie et al., 2019). In addition, *rps2* production is of great significance to the defense signal transduction process (Bent et al., 1994). Thus, in terms of genes coding confirmed products (*infA* & *rps*2), their pseudogenization can be used to adjust the transcription and signal transduction of *Rhizirideum* plants in response to the changing environment.

Table 4
Summary of pseudogenes and their productions in *Rhizirideum* plastomes

Pseudogene	Position	Species	Production
ycf68	IR	*A. polyrhizum, A. nutans, A. mongolicum*, *A. caespitosum, A. tenuissimum*, *A. anisopodium*	Putative protein RF68
orf56	IR	All	Putative protein RF56
ycf15	IR	All	Putative protein RF15
rps2	LSC	All	Ribosomal protein S2
infA	LSC	All	Trnaslation initiation factor 1

The SC/IR borders of angiosperm plastomes are generally conservative, lying mostly beside *rps19* and *ycf1* (Downie and Jansen, 2015). Genes *trnH-GUG* and *trnN-ACG* are believed to be located at the IR/LSC and IR/SSC borders of the ancestor of monocots, respectively (Zhu et al., 2016). This indicates an expansion of IR regions according to the relative position of *rps19* and *trnH-GUG*, as well as *ycf1* and *trnN-ACG* (Fig. 2). Generally speaking, the expansion of IRs can lead to the movement of SC/IR borders. Most terrestrial plants as subgen. *Rhizirideum* species present movements to a tiny extent, which can make a few genes into or out of IRs (Goulding et al., 1996; Wang et al., 2008; Wu and Chaw, 2015). Nonetheless, some plants do have their IRs expanding in a large scale. This can contribute to large increase or loss of IR genes, such as species in *Pelargonium*, *Psilotum*, *Leguminosae*, and *Erodium* (Palmer et al., 1987; Raubeson and Jansen, 1992; Tsudzuki et al., 1992; Chumley et al., 2006; Guisinger et al., 2010; Grewe et al., 2013; Sun et al., 2013; Guo et al., 2014). In *Rhizirideum* plastomes, the duplicated *rps19* moved into the IRs from the LSC, while the incompletely duplicated *ycf1* moved to cover the IR/SSC borders from the SSC (Fig. 5). In addition, the LSC/IRb boundaries also present a slight shift to the *rpl22* gene. The movements of IR/SC borders of subgen. *Rhizirideum* plastomes are tiny compared to the species mentioned above. Despite this, the IR expansion of our taxa is somewhat significant for the evolution. It is known that IR regions possess the nature of self-duplication, which has been proven to reduce the synonymous mutation rate (Ks) of genes, resulting in the Ks of IR genes being generally lower than that of SC genes [36]. It can be inferred that in the *Rhizirideum* plastomes, the Ks value of the *rps19* gene decreased after the movement from the LSC to the IRs. That is, the *rps19* gene has been more conserved, as well as its product, ribosomal protein S19, which is a component of the 40S ribosomal subunit. Therefore, it may contribute to the increace of stability of the ribosomal structure when *rps19* moved to IRs. This is also true of the gene *rpl22* coding ribosomal protein L22, a component of the 60S ribosomal subunit. The moving trend of *rpl22* may also influence the ribosomal structure. As is known, the structure of ribosomes can influence the expression pattern of genes, which are often relative with the environment (Ishihama, 2000). Consequently, the shift of IR/SC boundaries may be regarded as the adaptive evolution of plastomes. However, the product of *ycf1* and its functions should be studied further in the future.

There are twenty-six genes with introns in the plastome of *A. senescens*, three more than *Anena sativa* in the family *Gramineae* (single-copy gene *clp*, *rpoC*1 and double-copy gene *ycf*68). The transpliced gene *rps12* has three exons with one in the LSC and two in IRs. In eukaryotes, intron-splicing enhances gene expression by reducing transcription-associated mutagenesis (Niu and Yang, 2011). Meanwhile, this process imposes some selection pressure on genes (Petersen et al., 2011). Therefore, the intron-existing genes in *Rhizirideum* plastomes indicate that they are also under this kind of pressure.

The GC contents of *Rhizirideum* plastomes range from 36.8–36.9%, which is in accordance with those of many other monocots, approximately 37% (Huotari and Korpelainen, 2012; Liu et al., 2012; Peredo et al., 2013). Additionally, the GC content of *Allioideae* plastomes is below that of other taxa, such as Asparagaceae, Iridaceae, Agapanthaceae, etc. (Xie et al., 2020). This decrease can be attributed to the selective pressure caused by either neutral mutation (Ogata et al., 2001; Lane et al., 2007; Smith and Lee, 2008) or high transcription efficiency (Manen et al., 1998; And and Voelker, 2003). This is the same as the low GC content of the *Rhizirideum* plastomes.

3.2 Phylogenetic analysis

Phylogenetic analysis results demonstrate that subgen. *Rhizirideum* is a strongly supported monophyletic group, which corresponds with previous reports (Friesen et al., 2006). Species in section *Eduardii* (*A. przwalskianum* & *A. eduardii*) and section *Tenuissima* (*A. anisopodium* & *A. tenuissimum*) cluster into two individual branches, which is the same as the phylogenetic analysis of Li et al. (2016) based on ITS-rps16 data. Nevertheless, the interspecific relationships of the other two sections, *Rhizirideum* and *Rhizomatosa*, are somewhat different. The section *Rhizirideum* was divided into Asiatic and European geographical groups by Sinitsyna et al. (2016) and the species *A. senescens*, *A. spirale* and *A. nutans* were in the former group. However, the relationships among the *A. senescens*, *A. spirale* and *A. nutans* were not so clearly shown. Our results show that *A. senescens* and *A. spirale* form a sister branch and then cluster with *A. nutans* with a 100% support rate. Then Friesen et al. (2020) conducted a phylogenetic analysis for species in three *Rhizirideum* sections (*Caespitosoprason, Rhizomatosa*, and *Caespitosoprason*) based on chloroplast DNA fragments, where section *Caespitosoprason*, including *A. polyrhizum*, was merged into section *Rhizomatosa*, whereas our plastome tree indicates that *A. polyrhizum* is not clustered with *Rhizomatosa* species but with the remaining eleven species in subgenus *Rhizirideum*. In other words, *A. polyrhizum* should be separated from *Rhizimatosa* and placed into *Caespitosoprason*, which is believed to be a basal taxon of this subgenus here. Despite this, three species in section *Rhizomatosa* (*A. caespitosum*, *A. bidentatum* and *A. mongolicum*) are rather similar in morphology (leaf semiterete to terete and shorter than scape, perianth light red to purple, inner filament enlarged at base, bulbs densely clustered and the outer skin fibrous or striped broken). Disagreements between molecular and morphological analyses have also been frequently reported in other taxa, for instance, section *Daghestanica* and subgenus *Cyathophora* in genus *Allium* (Huang et al., 2014; Xie et al., 2019).

3.3 Subgen. Rhizirideum adaptability to the environment

The selective pressure analysis indicates that most of the common single-copy genes of subgen. *Rhizirideum* plastomes are under purifying selection, which also shows the conservation of the evolution of the plastome genes. There were eleven genes under relaxed selection ($0 < \omega < 0.5$), including *atpB*, *psbD*, and *rbcL* (Fig. 9, Table 3). Nine of those genes are involved in the phososystem I (e.g. *psaB*), photosystem II (e.g. *psbB*), and the electron transfer chain (e.g. *rbcL*). It is known that photosynthesis provides nutrients for plants and the electron transfer chain produces the energy substance, ATP (Weiss et al., 1991; Cramer et al., 2011; Xiao et al., 2012). And the PS II photoreaction center complex is mainly composed of proteins D1 and D2, among which D2 is coded by the *psbD* gene (Marder et al., 1987). Protein D2 plays an important role in both electron transfer and regulation of D1's expression (Diner and Rappaport, 2002). It has been found that *Rhizirideum* species are distributed mostly in the Eurasian stipa steppe with a temperate continental climate, which means very drought (Li et al., 2016; Sinitsyna et al., 2016). Therefore, most of the species of this subgenus prefer habitats of water enrichment such as pasture and shrub slopes, where there is greater sunlight competition pressure among all kinds of shrubs and herbs. The *psbD* gene in most fern lineages bears strong negative selection, while branches of tree-ferns obtain a high ω value, which may be related to the competition pressure of sunlight in the habitat (Page, 2002; Wang et al., 2012; Ke et al., 2013). This can also explain the relaxed selection of *psbD* in *Rhizirideum* plastomes. RuBisCo, encoded by the *rbcL* gene, is a major enzyme in photosynthesis. RuBisCo tends to evolve to enhance the affinity for CO_2 if the environment becomes hotter and drier (Liu et al., 2010). Thus, the *rbcL* gene of the subgenus *Rhizirideum* may bear the relaxed selection to affect the efficiency of photosynthesis. In addition, *rpl32* codes the ribosomal protein *L32*, and *rpoA* codes the RNA polymerase alpha subunit. RNA polymerase can regulate the pattern of gene expression under varying environments (Liu et al., 2010). Therefore, *rpl32* and *rpoA* may assist plants in adapting to extreme environments by influencing gene transcription and translation.

4. Methods

4.1 DNA isolation and sequencing

The fresh leaves of ten species were sampled from public areas and dried with silica gel afterward (locality see Additional file 4: Table S3). Total genomic DNA was isolated from silica-dried leaf tissues with a modified CTAB method. The voucher

specimens were deposited at the herbarium of Sichuan University (Chengdu, China)(voucher specimens: H11072607 (SZ), De-qing Huang; ZCJ20210821 (SZ), Chun-jing Zhou; FX2020081001 (SZ), Xiao Fu; FX2020080902 (SZ), Xiao Fu; H11072807 (SZ), De-qing Huang; H11070501 (SZ), De-qing Huang; FX2020081401 (SZ), Xiao Fu; FX2020081501 (SZ), Xiao Fu; FX2020081901 (SZ), Xiao Fu; ZCJ2012081910 (SZ), Chun-jing Zhou). DNA libraries were prepared and sequenced with the Illumina HiSeq 2500 platform with PE150 bp reads.

4.2 Synonymous codon usage bias

Complete chloroplast genomes were reconstructed by NOVOPlasty v2.6.2 (Dierckxsens et al., 2016) using *A. cepa* (MK335926) and *A. sativum* (MK335928) as references. Then the annotation of the plastid genomes was conducted with PGA (Qu et al., 2019) and manually adjusted with GENEIOUS R11 (Biomatters, Ltd., Auckland, New Zealand). Finally, the plastome circus map was drawn with OGDRAW (Greiner et al., 2019) and Gview (Petkau et al., 2010).

4.3 Sequence divergence

The program mVISTA (Frazer et al., 2004) was used to generate the whole-genome alignment of the eleven plastomes with A. senescens as a reference. Within GENEIOUS R11 all plastomes were aligned with MAFFT. The nucleotide diversity (Pi) of genes and intergenic regions was calculated by DNASP v.6 (Rozas et al., 2017).

4.4 repeat structure

REPuter (Stefan et al., 2001) was used to examine plastome repeat sequences. There were four matches of repeats classified: forward, palindromic, reverse, and complimentary matches. The parameters were as follows: repeat size of (1) > 30 bp; (2) > 90% sequence identity between the two repeats; and (3) Hamming distance = 3. Simple sequence repeats (SSRs) were identified by Perlscript MicroSAtellite (MISA). The setting motif sizes were one to six nucleotides, and the minimum repeat units were defined as 10, 5, 5, 4, 3 and 3 for mono-, di-, tri-, tetra-, penta- and hexa-nuleotides, respectively.

4.5 Phylogenetic analysis

In addition to ten plastomes newly sequenced, another twenty-three species were selected (including twelve Allium species from our team) (Additional file 3: Table S2) to infer the phylogenetic relationships. Seventy-one common single-copy genes were extracted from thirty-three taxa and were aligned. The alignments were trimmed and then concatenated in series for the phylogenetic analysis. The ML analysis was performed by RAxML v8.2.8 (Stamatakis, 2014) with the GTR + G model and 1000 bootstrap replicates. The BI analysis was performed by MrBayes v3.2.7 (Ronquist et al., 2012) with the substitution model GTR + I + Γ. The Markov chain Monte Carlo (MCMC) algorithm was run for one million generations, and one tree was sampled every 1000 generations. We then determined the MCMC convergence according to the average standard deviation of split frequencies (ASDSF) below 0.01. The first 20% of the trees were discarded as burn-in and the remaining trees were used to generate consensus trees. The non-coding sequences and the CDS sequences were also utilized to infer the phylogenetic relationships of the studied taxa.

4.6 Selective pressure analysis

We used an optimized branch-site model and Bayesian Emprical Bayes (BEB) methods (Yang et al., 2005) to identify genes under positive selection in species of our subgenus compared to those of other Allium taxa. Sixty-eight single-copy CDSs of thirty-three taxa were extracted and aligned with the software MUSCLE v5 (Edgar, 2021) aligned by codons. The positive selection analyses, measured by the ratio (ω) of the non-synonymous substitution rate (Ka) to the synonymous substitution rate (Ks), were performed using the branch-site model in the PAML v4.8 package (Yang, 2007), and our subgenus lineage was designated. Positive, neutral, and negative selection are demonstrated when the ratio $\omega > 1$, $w = 1$, and $\omega < 1$, respectively (Yang and Nielson, 2002). The log-likelihood values were tested (LRT) in accordance with (Lan et al., 2017). The BEB method was applied to compute the posterior probabilities of amino acid sites and those with a higher posterior probability were determined to be under positive selection. Sixty-eight single-copy sequences of 11 studied species were also used to calculate pairwise Ka/Ks ratios with KaKs Calculator v2.0 (Wang et al., 2010).

5. Conclusions

Our work revealed that (1) section *Caespitosoprason* including *Allium polyrhizum* should be separated from section *Rhizomatosa*, (2) section *Caespitosoprason* is a basal taxon of subgenus *Rhizirideum* and (3) plastome genes under relaxed selection contribute to the adaptability to the environment. Much still remains to be investigated on the phylogenetic relationships of species in subgenus *Rhizirideum*, notably improving the sampling of other species of this genus.

Declarations

Ethical approval and consent to participate

Plant material complies with local and national regulations. The voucher specimens were deposited at the herbarium of Sichuan University (Chengdu, China)(voucher specimens: H11072607 (SZ), De-qing Huang; ZCJ20210821 (SZ), Chun-jing Zhou; FX2020081001 (SZ), Xiao Fu; FX2020080902 (SZ), Xiao Fu; H11072807 (SZ), De-qing Huang; H11070501 (SZ), De-qing Huang; FX2020081401 (SZ), Xiao Fu; FX2020081501 (SZ), Xiao Fu; FX2020081901 (SZ), Xiao Fu; ZCJ2012081910 (SZ), Chun-jing Zhou).

Consent for publication

Not applicable.

Conflict of interest

The authors declare that they have no competing interests.

Author contribution

Xiao Fu: Conceptualization, Resources, Data curation, Formal analysis, Writing - original draft, Writing - review & editing. **Deng-Feng Xie:** Data curation, Writing - review & editing. **Yu-Yang Zhou:** Resources, Data curation. **Rui-Yu Cheng:** Resources, Writing - review & editing. **Xiang-Yi Zhang:** Resources, Writing - review & editing. **Xing-Jin He:** Resources, Data curation, Writing - review & editing. **Song-Dong Zhou:** Resources, Writing - review & editing.

Acknowledgement

We would like to acknowledge all the professors and fellow students in the laboratory.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files. The datasets analyzed during the current study are available in the NCBI GenBank repository (See supplementary Additional file 3: Table S2 for accessions).

Funding

This work was supported by the National Natural Science Foundation of China (Grant Nos. 32100180, 32070221, 32170209), the Fundamental Research Funds for the Central Universities (20826041E4158), the China Postdoctoral Science Foundation (2020M683303).

References

1. And, K.D., and Voelker, L.R.L. (2003). Molecular biology of mycoplasmas. *Annu. Rev. Microbiol.* 50, 25.

2. Bent, A.F., Kunkel, B.N., Dahlbeck, D., Brown, K.L., and Schmidt, R., et al. (1994). RPS2 of *Arabidopsis thaliana*: A leucine-rich repeat class of plant disease resistance genes. *Science* 265, 1856-1860. https://doi.org/10.1126/science.8091210
3. Chumley, T.W., Palmer, J.D., Mower, J.P., Matthew, F.H., and Calie, P.J., et al. (2006). The complete chloroplast genome sequence of pelargonium × hortorum: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol.*, 2175-2190. https://doi.org/10.1093/molbev/msl089
4. Cramer, W.A., Yamashita, E., Baniulis, D., and Hasan, S.S. (2011). *The Cytochrome b6f Complex of Oxygenic Photosynthesis*. Encyclopedia of Inorganic and Bioinorganic Chemistry.
5. Dierckxsens, N., Mardulyn, P., and Smits, G. (2016). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 4, e18. https://doi.org/ 10.1093/nar/gkw955
6. Diner, B.A., and Rappaport, F. (2002). Structure, dynamics, and energetics of the primary photochemistry of photosystem ii of oxygenic photosynthesis. *Annu. Rev. Plant Biol.* https://doi.org/10.1146/annurev.arplant.53.100301.135238
7. Do, G.S., and Seo, B.B. (2000). Phyiogenetic relationships among Allium subg. Rhizirideum species based on the molecular variation of 5S rRNA genes. *Kjbs* 4, 77-85.
8. Downie, S.R., and Jansen, R.K. (2015). A comparative analysis of whole plastid genomes from the apiales: Expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Syst. Bot.* 40, 336-351. https://doi.org/10.1600/036364415X686620
9. Dubouzet, J., Shinoda, K., and Murata, N. (1997). Phylogeny of Allium L. Subgenus Rhizirideum (G. Don ex Koch) Wendelbo according to dot blot hybridization with randomly amplified DNA probes. *Theor. Appl.* 95, 1223-1228. https://doi.org/10.1007/s001220050685
10. Edgar, R. (2021). MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. *BioRxiv*. https://doi.org/10.1101/2021.06.20.449169
11. Frazer, K.A., Lior, P., Alexander, P., Rubin, E.M., and Inna, D. (2004). VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* 32(Web Server issue), W273-W279. https://doi.org/10.1093/nar/gkh458
12. Friesen, N. (1992). Systematics of the Siberian polyploid complex in subgenus Rhizirideum (Allium). *Hanelt P, Hammer K. And Knüpffer (Eds.) the Genus Allium - Taxonomie Problemms and Genetic Resources. Proceedings of an International Symposium Held at Gatersleben. Germany. June 11-13. 1991. Pp. 55-66.*
13. Friesen, N., Fritsch, R.M., and Blattner, F.R. (2006). Phylogeny and new intrageneric classification of *allium* (Alliaceae) based on nuclear ribosomal DNA its sequences. *Aliso* 22, 372-395. https://doi.org/10.5642/aliso.20062201.31
14. Friesen, N., Smirnov, S., Shmakov, A., Herden, T., and Batlai, O., et al. (2020). Allium species of section Rhizomatosa, early members of the Central Asian steppe vegetation. *Flora* 263, 151536. https://doi.org/10.1016/j.flora.2019.151536
15. Goulding, S.E., Wolfe, K.H., Olmstead, R.G., and Morden, C.W. (1996). Ebb and flow of the chloroplast inverted repeat. *Molecular and General Genetics MGG* 252, 195-206. https://doi.org/Mol. Gen. Genet.
16. Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59-W64. https://doi.org/10.1093/nar/gkz238
17. Grewe, F., Guo, W., Gubbels, E., Hansen, A., and Mower, J. (2013). Complete plastid genomes from Ophioglossum californicum, Psilotum nudum, and Equisetum hyemale reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. *BMC Evol. Biol.* 13. https://doi.org/10.1186/1471-2148-13-8
18. Guisinger, M.M., Kuehl, J.V., Boore, J.L., and Jansen, R.K. (2010). Extreme reconfiguration of plastid genomes in the angiosperm family geraniaceae: Rearrangements, repeats, and codon usage. *Mol. Biol. Evol.* https://doi.org/10.1093/molbev/msq229
19. Guo, W., Felix, G., Amie, C.C., Fan, W., and Duan, Z., et al. (2014). Predominant and substoichiometric isomers of the plastid genome coexist within juniperus plants and have shifted multiple times during cupressophyte evolution. *Genome Biol. Evol.* 6, 580-590. https://doi.org/10.1093/gbe/evu046

20. Herden, T., Hanelt, P., and Friesen, N. (2016). Phylogeny of Allium L. Subgenus Anguinum (G. Don. Ex W.D.J. Koch) N. Friesen (Amaryllidaceae). *Mol. Phylogenetics Evol.* 95, 79-93. https://doi.org/10.1016/j.ympev.2015.11.004
21. Huang, D.Q., Yang, J.T., Zhou, C.J., Zhou, S.D., and He, X.J. (2014). Phylogenetic reappraisal of Allium subgenus Cyathophora (Amaryllidaceae) and related taxa, with a proposal of two new sections. *J. Med. Plant Res.* 127, 275-286. https://doi.org/10.1007/s10265-013-0617-8
22. Huang, Y., Cho, S., Haryono, M., and Kuo, C. (2017). Complete chloroplast genome sequence of common bermudagrass (*Cynodon dactylon* (L.) Pers.) and comparative analysis within the family Poaceae. *PLoS ONE* 12, e179055. https://doi.org/10.1371/journal.pone.0179055
23. Huotari, T., and Korpelainen, H. (2012). Complete chloroplast genome sequence of *Elodea canadensis* and comparative analyses with other monocot plastid genomes. *Gene* 508, 96-105. https://doi.org/10.1016/j.gene.2012.07.020
24. Ishihama, A. (2000). Functional modulation of escherichia coli RNA polymerase. *Annual Review of Microbiology* 54, 499-518. https://doi.org/10.1146/annurev.micro.54.1.499
25. Jang, J.E., Park, J.S., Jung, J.Y., Kim, D.K., and Yang, S., et al. (2021). Notes on Allium section Rhizirideum (Amaryllidaceae) in South Korea and northeastern China: With a new species from Ulleungdo Island. *PhytoKeys* 176, 1-19. https://doi.org/10.3897/phytokeys.176.63378
26. Jin, G.H., Chen, S.Y., Ting-Shuang, Y.I., and Zhang, S.D. (2014). Characterization of the complete chloroplast genome of apple (Malus × domestica,rosaceae). *Plant Diversity and Resources*, 468-484.
27. Kamelin, P.B. (1973). *Florogeneticheskij analiz estestvennoj flory gornoj Srednej Azii*. Nauka: Leningrad, Russia.
28. Ke, X.U., Wang, B., Ying-Juan, S.U., Gao, L., and Wang, T. (2013). Molecular evolution of d gene in ferns:selection pressure and co-evolutionary analysis. *Plant Sci. J.* 31, 429. https://doi.org/10.3724/SP.J.1142.2013.50429
29. Kim, H.H., Kang, H.W., Park, Y.J., Baek, H.J., and Gwag, J.K. (2001). Phylogenic relationship of allium species in subgenus rhizirideum by PCR DNA fingerprint. *Korean J. Crop Sci.* 46, 328-333.
30. Lan, Y., Sun, J., Tian, R., Bartlett, D.H., and Li, R., et al. (2017). Molecular adaptation in the world's deepest-living animal: Insights from transcriptome sequencing of the hadal amphipod *Hirondellea gigas*. *Mol. Ecol.* 26, 3732-3743.
31. Lane, C.E., van den Heuvel, K., Kozera, C., Curtis, B.A., and Parsons, B.J., et al. (2007). Nucleomorph genome of *Hemiselmis andersenii* reveals complete intron loss and compaction as a driver of protein structure and function. *Proc Natl Acad Sci U S a* 104, 19908-19913. https://doi.org/10.1073/pnas.0707419104
32. Lee, H.L., Jansen, R.K., Chumley, T.W., and Kim, K.J. (2007). Gene relocations within chloroplast genomes of jasminum and menodora (Oleaceae) are due to multiple, overlapping inversions. *Mol. Biol. Evol.* 24, 1161-1180. https://doi.org/10.1093/molbev/msm036
33. Lee, N.S. (2001). Phylogenetic analyses of nuclear rDNA its sequences of Korean Allium L. Subgenus Rhizirideum (Alliaceae). *Kjbs* 5, 283-290.
34. Li, L., Hu, Y., He, M., Zhang, B., and Wu, W., et al. (2021). Comparative chloroplast genomes: Insights into the evolution of the chloroplast genome of Camellia sinensis and the phylogeny of Camellia. *BMC Genom.* 22.
35. Li, Q.Q., Zhou, S.D., Huang, D.Q., He, X.J., and Wei, X.Q. (2016). Molecular phylogeny, divergence time estimates and historical biogeography within one of the world's largest monocot genera. *AoB Plants* 8, w41. https://doi.org/10.1093/aobpla/plw041
36. Linnaeus, C. (1753). *Species plantarum: Exhibentes plantas rite cognitas ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas*.
37. Liu, J., Qi, Z.C., Zhao, Y.P., Fu, C.X., and Jenny, X.Q. (2012). Complete cpDNA genome sequence of *Smilax china* and phylogenetic placement of Liliales–influences of gene partitions and taxon sampling. *Mol. Phylogenet. Evol.* 64, 545-562. https://doi.org/10.1016/j.ympev.2012.05.010
38. Liu, N., Wang, Q., Chen, J., Zhu, Y., and Tashi, T., et al. (2010). Adaptive evolution and structure modeling of rbcL gene in Ephedra. *Chin. Sci. Bull.* 55, 2341-2346. https://doi.org/10.1007/s11434-010-3023-9

39. Liu, Q., Li, X., Li, M., Xu, W., and Heslop-Harrison, J.S. (2020). Comparative chloroplast genome analyses of *Avena*: Insights into evolutionary dynamics and phylogeny. *BMC Plant Biol.* 20. https://doi.org/10.1186/s12870-020-02621-y

40. Manen, J.F., Cuénoud, P., and Martinez, M.D.P. (1998). Intralineage variation in the pattern of *rbcL* nucleotide substitution. *Plant Sys. Evol.* 211, 103-112.

41. Marder, J.B., Chapman, D.J., Telfer, A., Nixon, P.J., and Barber, J. (1987). Identification of psbA and psbD gene products, D1 and D2, as reaction centre proteins of photosystem 2. *Plant Mol. Biol.* 9, 325-333.

42. McKain, M.R., McNeal, J.R., Kellar, P.R., Eguiarte, L.E., and Pires, J.C., et al. (2016). Timing of rapid diversification and convergent origins of active pollination within Agavoideae (Asparagaceae). *Am. J. Bot.* 103, 1717-1729. https://doi.org/10.3732/ajb.1600198

43. Niu, D.K., and Yang, Y.F. (2011). Why eukaryotic cells use introns to enhance gene expression: Splicing reduces transcription-associated mutagenesis by inhibiting topoisomerase I cutting activity. *Biol. Direct* 6, 24. https://doi.org/10.1186/1745-6150-6-24

44. Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.E., and Barbe, V., et al. (2001). Mechanisms of evolution in Rickettsia conorii and R. Prowazekii. *Science* 293, 2093-2098. https://doi.org/10.1126/science.1061471

45. Özler, H., and Pehlivan, S. (2010). Pollen morphology of some *Allium* L. (Lilliaceae) taxa in turkey. *Bangladesh J. Bot.* 39. https://doi.org/10.3329/bjb.v39i1.5524

46. Page, C.N. (2002). Ecological strategies in fern evolution: A neopteridological overview. *Rev. Palaeobot. Palynol.* 119, 1-33. https://doi.org/10.1016/S0034-6667(01)00127-0

47. Palmer, J.D., Osorio, B., Aldrich, J., and Thompson, W.F. (1987). Chloroplast DNA evolution among legumes: Loss of a large inverted repeat occurred prior to other sequence rearrangements. *Curr. Genet.* 11, 275-286. https://doi.org/10.1007/BF00355401

48. Park, S., Jansen, R.K., and Park, S. (2015). Complete plastome sequence of Thalictrum coreanum (Ranunculaceae) and transfer of the rpl32 gene to the nucleus in the ancestor of the subfamily Thalictroideae. *BMC Plant Biol.* 15, 40. https://doi.org/10.1186/s12870-015-0432-6

49. Peredo, E.L., King, U.M., and Les, D.H. (2013). The plastid genome of *najas flexilis*: Adaptation to submersed environments is accompanied by the complete loss of the NDH complex in an aquatic angiosperm. *Plos One* 8, e68591. https://doi.org/10.1371/journal.pone.0068591

50. Petersen, K., Schottler, M.A., Karcher, D., Thiele, W., and Bock, R. (2011). Elimination of a group II intron from a plastid gene causes a mutant phenotype. *Nucleic Acids Res.* 39, 5181-5192. https://doi.org/10.1093/nar/gkr105

51. Petkau, A., Stuart-Edwards, M., Stothard, P., Domselaar, G.V., and Valencia, A. (2010). Interactive microbial genome visualization with GView. *Bioinformatics* 26, 3125-3126. https://doi.org/10.1093/bioinformatics/btq588

52. Qu, X.J., Moore, M.J., Li, D.Z., and Yi, T.S. (2019). PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* 15, 1-12. https://doi.org/10.1186/s13007-019-0435-7

53. Raamsdonk, L., Ginkel, V.V., and Kik, C. (2000). Phylogeny reconstruction and hybrid analysis in Allium subgenus Rhizirideum. *Theor. Appl.* 100, 1000-1009.

54. Raamsdonk, L.V., Ensink, W., Heusden, A., Ginkel, M., and Kik, C. (2003). Biodiversity assessment based on cpDNA and crossability analysis in selected species of Allium subgenus Rhizirideum. *Theor. Appl.* 107, 1048-1058. https://doi.org/10.1007/s00122-003-1335-8

55. Raubeson, L.A., and Jansen, R.K. (1992). A rare chloroplast-DNA structural mutation is shared by all conifers. *Biochem. Syst. Ecol.* 20, 17-24. https://doi.org/10.1016/0305-1978(92)90067-N

56. Ravi, V., Khurana, J.P., Tyagi, A.K., and Khurana, P. (2006). The chloroplast genome of mulberry: Complete nucleotide sequence, gene organization and comparative analysis. *Tree Genet. Genomes* 3, 49-59. https://doi.org/10.1007/s11295-006-0051-3

57. Regel, E. (1875). Alliorum adhuc cognitorum monographia. *Acta Hort. Petropol.* 3, 1-266.

58. Regel, E. (1887). Allii species Asiae Centralis in Asia Media a Turcomania desertisque Araliensibus et Caspicis usque ad Mongolian crescentes. *Acta Hort. Petropol.* 10, 278-362.

59. Ren, T., Li, Z.X., Xie, D.F., Gui, L.J., and Peng, C., et al. (2020). Plastomes of eight *Ligusticum* species: Characterization, genome evolution, and phylogenetic relationships. *BMC Plant Biol.* 20, 519. https://doi.org/10.1186/s12870-020-02696-7

60. Ricroch, A., Yockteng, R., Brown, S., and Nadot, S. (2005). Evolution of genome size across some cultivated *Allium* species. *Genome* 48, 511-520. https://doi.org/10.1139/g05-017

61. Rola, K. (2014). Cell pattern and ultrasculpture of bulb tunics of selected *allium* species (Amaryllidaceae), and their diagnostic value. *Acta Biol. Crac. Ser. Bot.* 0. https://doi.org/10.2478/abcsb-2014-0002

62. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D., and Darling, A., et al. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Sys. Biol.* 61, 539-542. https://doi.org/10.1093/sysbio/sys029

63. Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., and Librado, P., et al. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. *Mol. Biol. Evol.* 34, 3299-3302. https://doi.org/10.1093/molbev/msx248

64. Shopova, M. (1966). The nature and behaviour of supernumerary chromosomes in the Rhizirideum group of the genus Allium. *Chromosoma* 19, 149-158. https://doi.org/10.1007/BF00293680

65. Sinitsyna, T., Herden, T., and Friesen, N. (2016). Dated phylogeny and biogeography of the Eurasian Allium section Rhizirideum (Amaryllidaceae). *Plant Syst. Evol.* 302. https://doi.org/10.1007/s00606-016-1333-3

66. Smith, D.R., and Lee, R.W. (2008). Mitochondrial genome of the colorless green alga *Polytomella capuana*: A linear molecule with an unprecedented GC content. *Mol Biol Evol* 25, 487-496. https://doi.org/10.1093/molbev/msm245

67. Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30. https://doi.org/10.1093/bioinformatics/btu033

68. Stefan, K., Choudhuri, J.V., Enno, O., Chris, S., and Jens, S., et al. (2001). REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 22, 4633-4642. https://doi.org/10.1093/nar/29.22.4633

69. Sun, Y.X., Moore, M.J., Meng, A.P., Soltis, P.S., and Soltis, D.E., et al. (2013). Complete plastid genome sequencing of trochodendraceae reveals a significant expansion of the inverted repeat and suggests a paleogene divergence between the two extant species. *Plos One* 8, e60429. https://doi.org/10.1371/journal.pone.0060429

70. Tang, D., Wei, F., and Zhou, R. (2021). Comparative analysis of chloroplast genomes of *Kenaf cytoplasmic* male sterile line and its maintainer line. *Sci. Rep.* 11.

71. Traub, H.P. (1968). The subgenera, sections and subsections of *Allium* L. *Plant Life* 24.

72. Tsudzuki, J., Nakashima, K., Tsudzuki, T., Hiratsuka, J., and Shibata, M., et al. (1992). Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: Nucleotide sequences of trnQ, trnK, psbA, trnI and trnH and the absence of rps16. *Mol. Gen. Genet.* 232, 206-214. https://doi.org/10.1007/BF00279998

73. Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. *Genomics Proteomics Bioinformatics* 1, 77-80. https://doi.org/10.1016/S1672-0229(10)60008-3

74. Wang, R.J., Cheng, C.L., Chang, C.C., Wu, C.L., and Su, T.M., et al. (2008). Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol. Biol.* 8. https://doi.org/10.1186/1471-2148-8-36

75. Wang, T., Su, Y., and Li, Y. (2012). Population genetic variation in the tree fern *alsophila spinulosa* (Cyatheaceae): Effects of reproductive strategy. *Plos One* 7, e41780. https://doi.org/10.1371/journal.pone.0041780

76. Weiss, H., Friedrich, T., Hofhaus, G., and Preis, D. (1991). The respiratory-chain NADH dehydrogenase (complex I) of mitochondria. *European Journal of Biochemistry* 197, 563-576.

77. Wen, F., Wu, X., Li, T., Jia, M., and Liu, X., et al. (2021). The complete chloroplast genome of *Stauntonia chinensis* and compared analysis revealed adaptive evolution of subfamily Lardizabaloideae species in China. *BMC Genom.* 22.
78. Wendelbo, P. (1969). New subgenera, sections and species of *Allium*. *Bot. Notiser* 122.
79. Wicke, S., Schneeweiss, G.M., De Pamphilis, C.W., Kai, F.M., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273-297. https://doi.org/10.1007/s11103-011-9762-4
80. Wu, C., and Chaw, S. (2015). Evolutionary stasis in cycad plastomes and the first case of plastome GC-Biased gene conversion. *Genome Biol. Evol.*, 2000-2009.
81. Xiao, J., Li, J., Ouyang, M., Yun, T., and He, B., et al. (2012). DAC is involved in the accumulation of the cytochrome b6/f complex in *Arabidopsis*. *Plant Physiology* 160, 1911-1922.
82. Xie, D., Tan, J., Yu, Y., Gui, L., and Su, D., et al. (2020). Insights into phylogeny, age and evolution of *Allium* (Amaryllidaceae) based on the whole plastome sequences. *Annals of Botany*, 7.
83. Xie, D.F., Huan-Xi, Y.U., Price, M., Xie, C., and He, X.J. (2019). Phylogeny of chinese allium species in section daghestanica and adaptive evolution of allium (Amaryllidaceae, allioideae) species revealed by the chloroplast complete genome. *Front. Plant Sci.* 10. https://doi.org/10.3389/fpls.2019.00460
84. Xie, D.F., Yu, Y., Wen, J., Huang, J., and He, X.J. (2020). Phylogeny and highland adaptation of Chinese species in Allium section Daghestanica (Amaryllidaceae) revealed by transcriptome sequencing. *Molecular Phylogenetics and Evolution* 146, 106737.
85. Yang, X., Xie, D.F., Chen, J.P., Zhou, S.D., and He, X.J. (2020). Comparative analysis of the complete chloroplast genomes in *allium* subgenus cyathophora (Amaryllidaceae): Phylogenetic relationship and adaptive evolution. *BioMed Res. Int.* 25, 1-17. https://doi.org/10.1155/2020/1732586
86. Yang, Z. (2007). PAML 4: A program package for phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*, 1586-1591.
87. Yang, Z., Wong, W.S.W., and Nielson, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.*, 1107-1118. https://doi.org/10.1093/molbev/msi097
88. Yang, Z., and Nielson, R. (2002). Codon-Substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.*, 908-917. https://doi.org/10.1093/oxfordjournals.molbev.a004148
89. Zhai, W., Duan, X., Zhang, R., Guo, C., and Li, L., et al. (2019). Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. *Mol. Phylogenet. Evol.* 135, 12-21. https://doi.org/10.1016/j.ympev.2019.02.024
90. Zhu, A., Guo, W., Gupta, S., Fan, W., and Mower, J. (2016). Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. *New Phytol.* 209, 1747-1756. https://doi.org/10.1111/nph.13743
91. Zyab, C., Tao, D.A., Sv, A., Fk, B., and Dmab, C., et al. (2021). Phylogenomics of Allium section Cepa (Amaryllidaceae) provides new insights on domestication of onion - ScienceDirect. *Plant Divers.* 43, 102-110. https://doi.org/10.1016/j.pld.2020.07.008

Figures

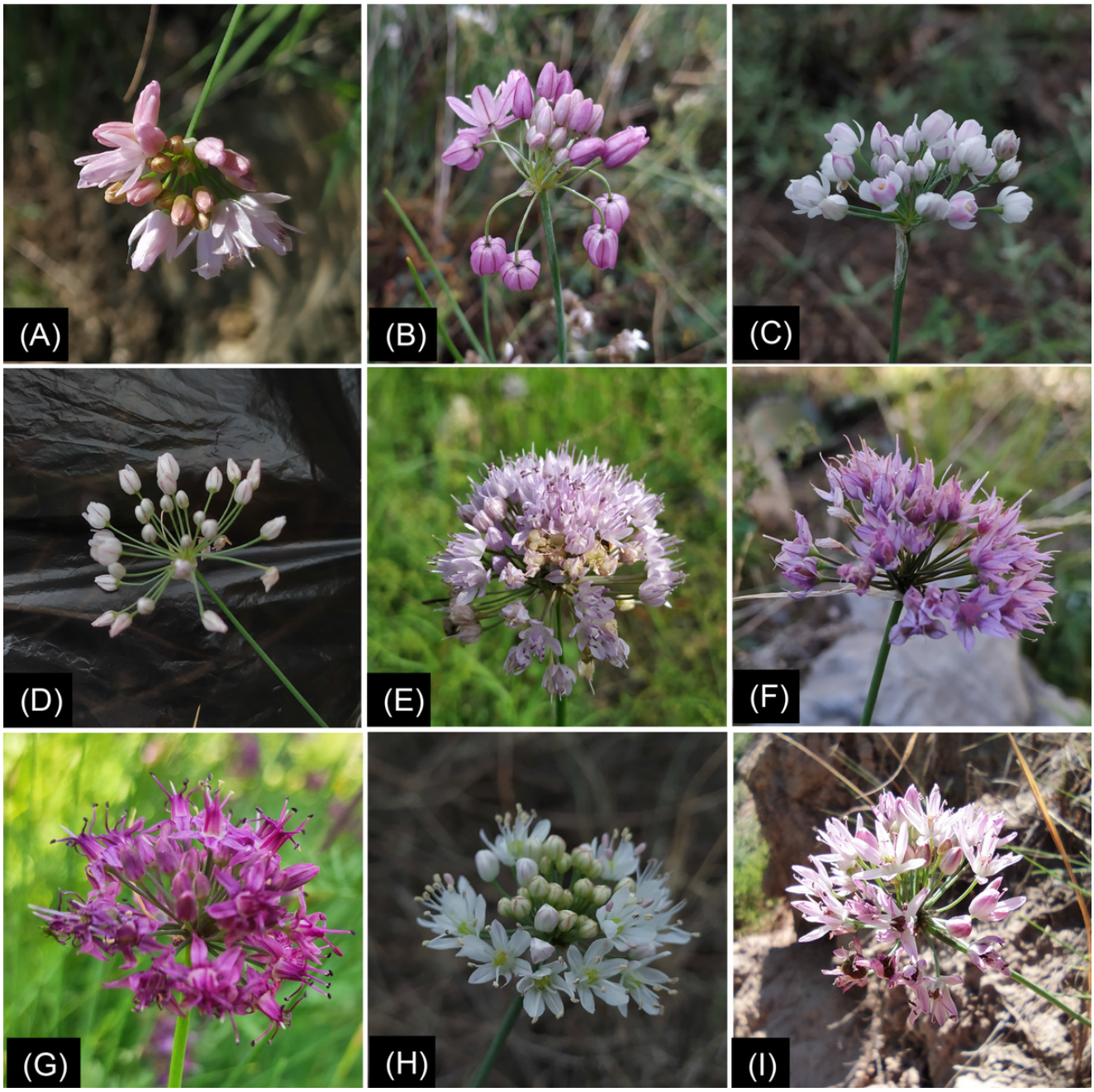

Figure 1

Inflorescenes of eight species in subgenus *Rhizirideum*. (A), *A.bidentatum*; (B), *A. mongolicum*; (C), *A. anisopodium*; (D), *A. tenuissimum*; (E), *A. senescens*; (F), *A. eduardii*; (G), *A. przewalskianum*; (H) & (I), *A. polyrhizum*.

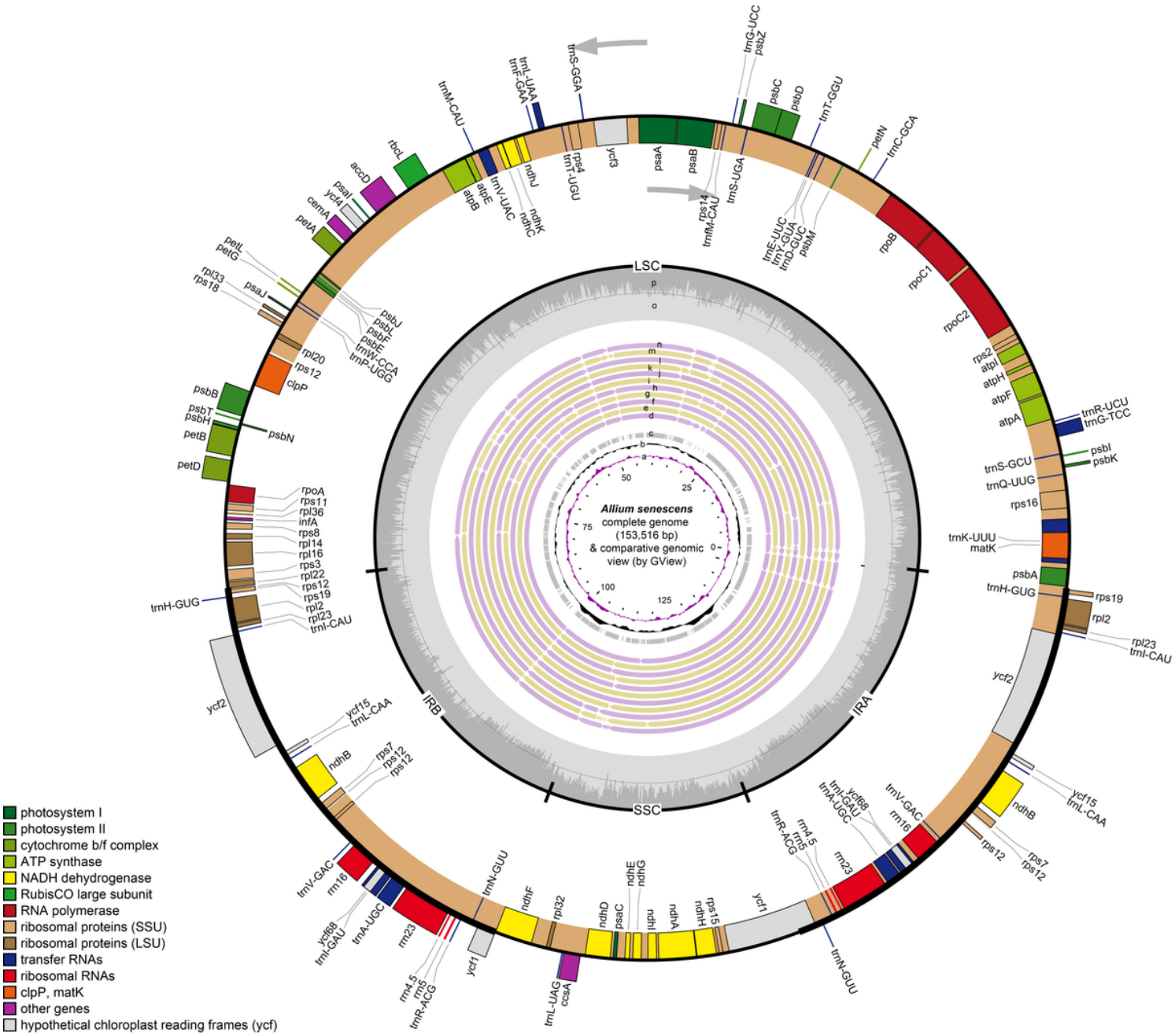

Figure 2

Chloroplast map of *Allium senescens* (the outermost circle and rings o-p) and GView comparison of eleven Rhizirideum plastomes (rings d-n). Genes are shown in different colors based on the functional groups they belong to. genes on the inside of the outer circle are transcribed clockwise and on the outside counter-clockwise. LSC, large single-copy region; SSC, small single-copy region; IR, inverted repeat. Rings a-b: GC skews and GC content deviations from *A. senescens* plastome GC content, respectively. Ring c: reference of multiple alignment (A. senescens plastome). Ring d-n denote the plastome sequence comparison by BLAST between A. senescens and other species plastomes outwards in turn: *A. bidentatum, A. caespitosum, A. mongolicum, A. anisopodium, A. tenuissimum, A. senescens, A. spirale, A. nutans, A. eduardii, A. przewalskianum, A. polyrhizum.* Rings o & p: AT and GC content of *A. senescens* plastome by OGDRAW.

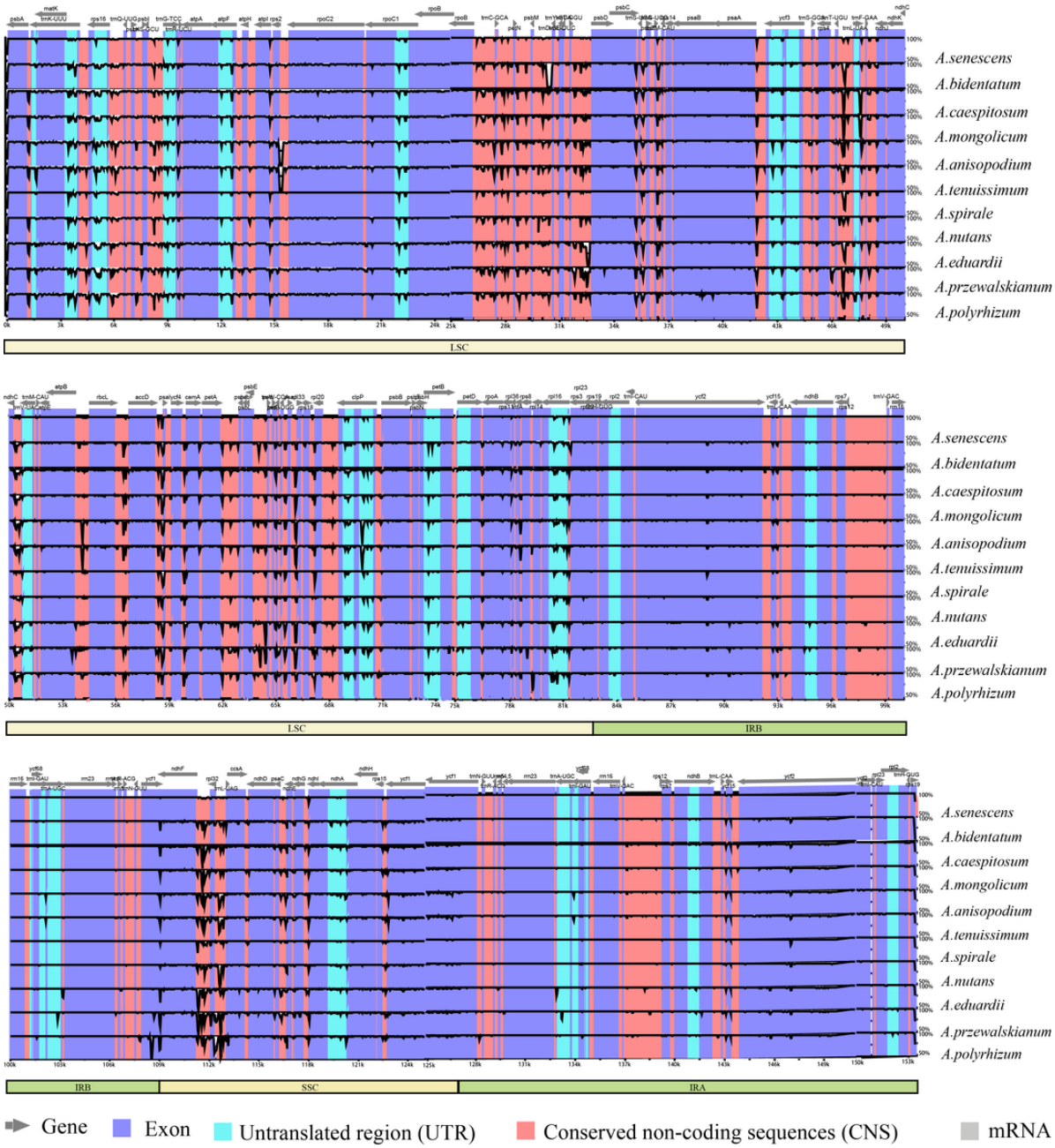

Figure 3

mVISTA comparison of eleven subgenus *Rhizirideum* plastomes (*A. senescens* as reference). Sequence regions of different types are shown in different colors.

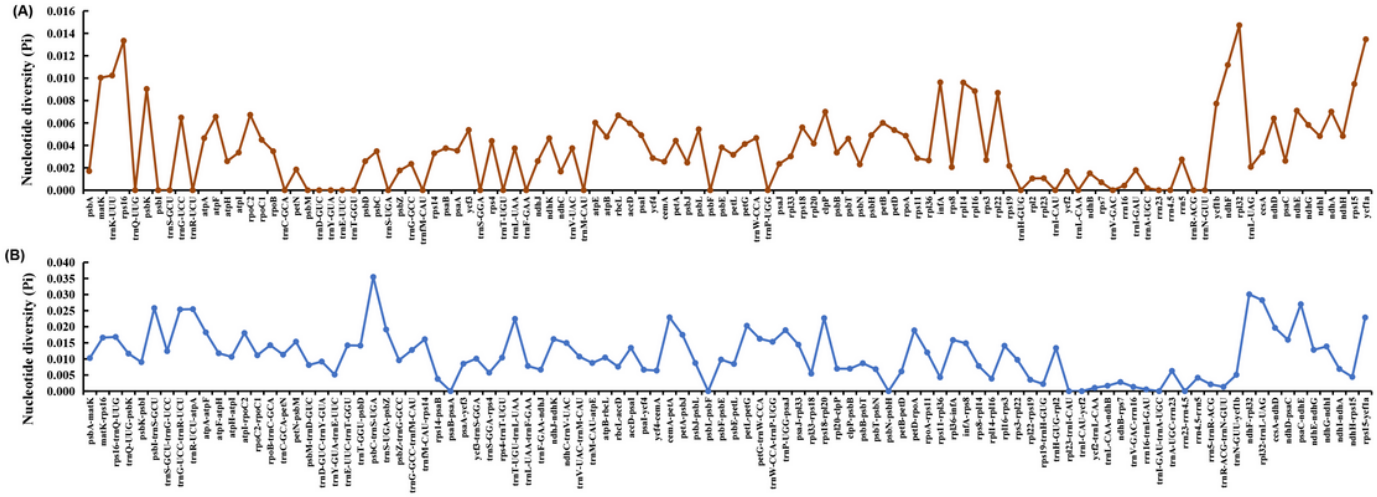
(A)
Nucleotide diversity (Pi)
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
(B)
Nucleotide diversity (Pi)
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000

Figure 4

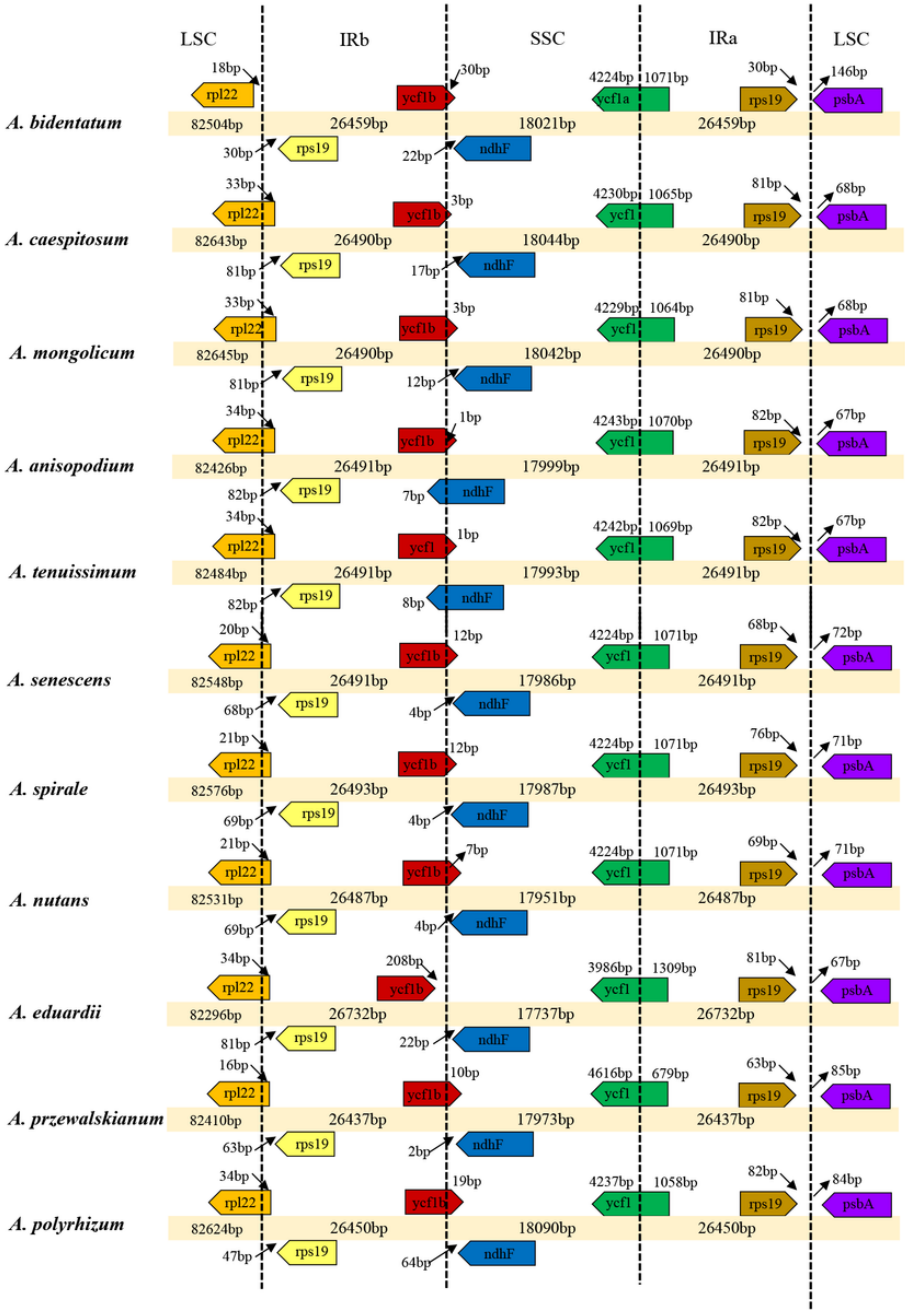

Figure 5

IR/SC boundaries of eleven subgenus *Rhizirideum* plastomes.

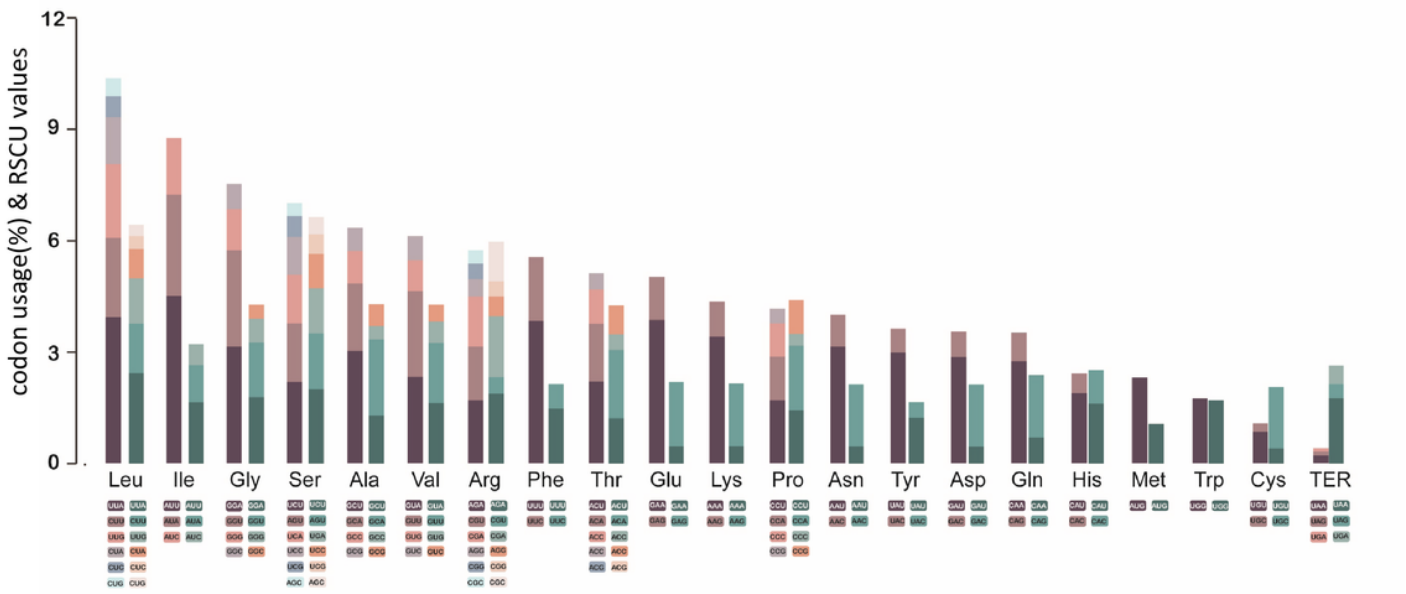

Figure 6

Codon usage in percentages (the left column) and RSCU values (the right column) of twenty-one amino acids. Each codon for an amino acid is shown with different colors.

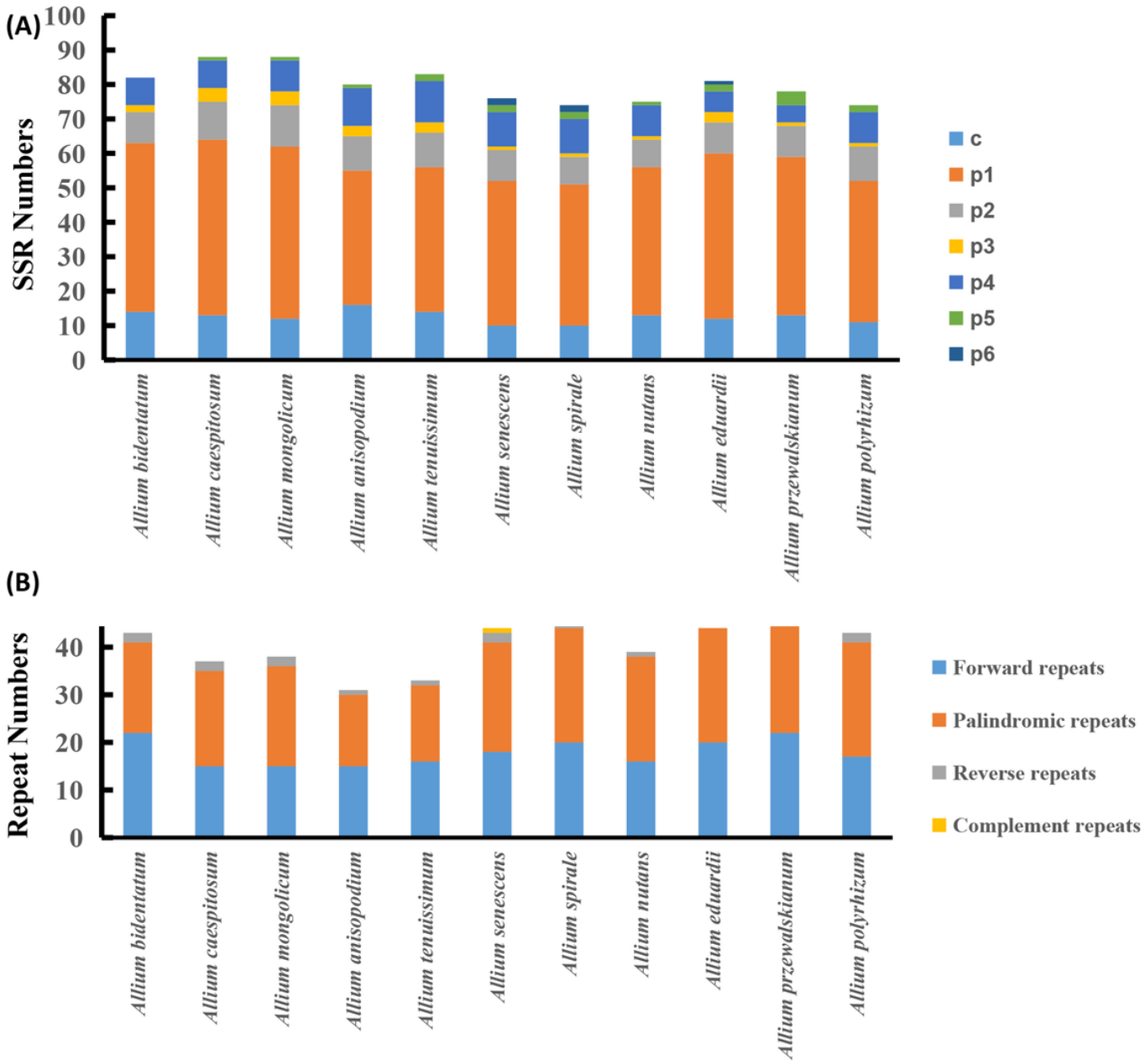

Figure 7

Numbers of SSRs and repeats of eleven *Rhizirideum* species. (A), stacking histogram of SSR numbers. Seven sorts of SSRs are shown with different colors: c, compound microsatellits; p1-p6, microsatellites with one to six bases as a repeat unit. (B), stacking histogram of repeats (30-50 bp) numbers. Four sorts of repeats are shown with different colors.

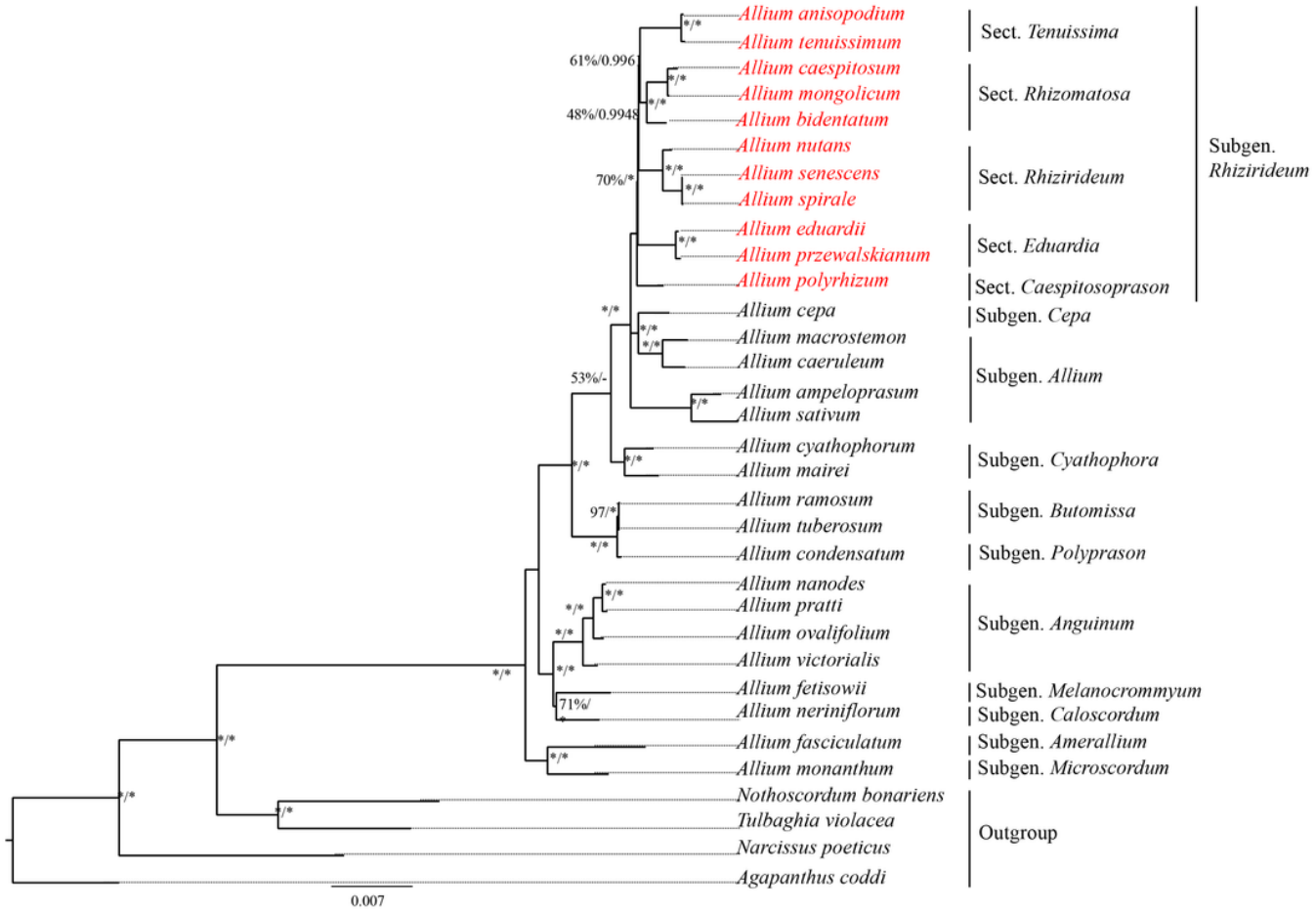

Figure 8

Phylogenetic tree reconstruction of 33 species inferred from Maximum likelihood (ML) and Bayesian inference (BI) analyses based on single-copy genes of plastomes. The bootstrap support values and posterior probability values are listed at each node. The asterisk (*) means 100% or 1. The minus sign (-) means no value.

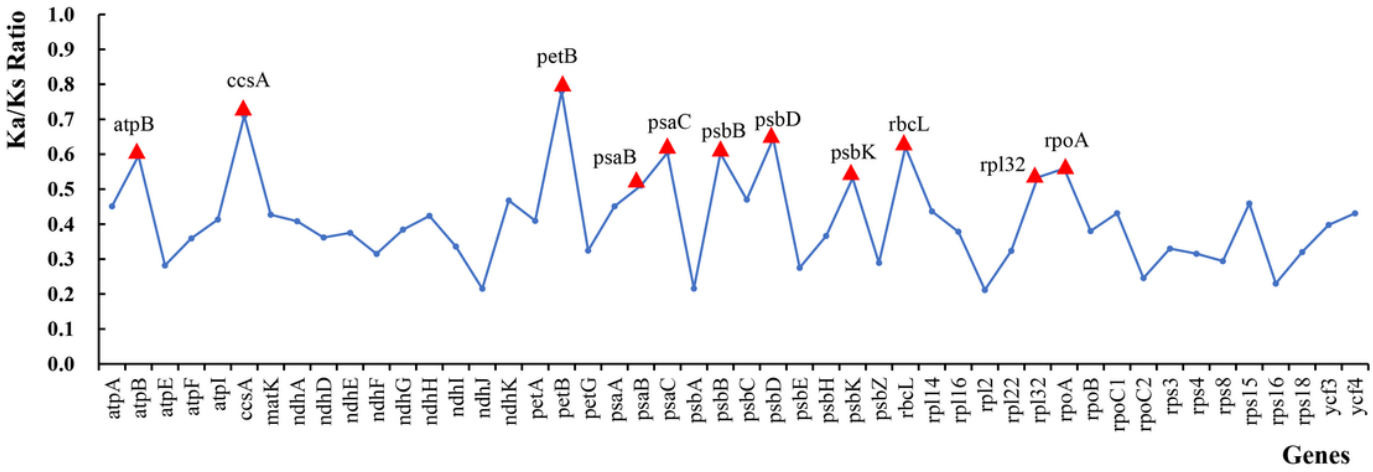

Figure 9

Ka/Ks ratios of 48 single-copy genes. The top eleven genes are noted with red triangles.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1FigureS1.Bulbshapesof9species.pdf
- Additionalfile2TableS1.ListofcommonsinglecopygenesinelevenRhizirideumplastomesforphylogeneticreconstrucion.docx
- Additionalfile3TableS2.ListofspeciesandtheiraccessionnumbersinGenBankincludedinthephylogeneticanalysis.docx
- Additionalfile4TableS3.Collectionlocalityandvoucherinformationof10sequencedplastomes.docx
- Additionalfile5TableS4S5.CodonusageandRSCUvaluesofproteincodinggenesoftheelevenRhizirideumplastomes.docx
- Additionalfile6TableS6.BasecompositionofproteincodinggenesofelevenRhizirideumplastomes.docx
- Additionalfile7TableS7S8.TherepeatsequenceandSSRdistributionintheelevenRhizirideumplastomes.docx
- Additionalfile8TabelS9.ResultsofselectivepressureanalysisinPAMLwiththebranchsitemodel.docx