Original article

Characterization of the complete chloroplast genome of *Zephyranthes phycelloides* (*Amaryllidaceae,* tribe *Hippeastreae*) from Atacama region of Chile

Roberto Contreras-Díaz [a,*], Mariana Arias-Aburto [a], Liesbeth van den Brink [b]

[a] *Centro Regional de Investigación y Desarrollo Sustentable de Atacama (CRIDESAT), Universidad de Atacama, Av. Copayapu 485, 1530000 Copiapó, Chile*
[b] *Department of Evolution and Ecology, Plant Ecology Group, Universität Tübingen, 72076 Tübingen, Germany*

A B S T R A C T

Sporadic rains in the Atacama Desert reveal a high biodiversity of plant species that only occur there. One of these rare species is the "Red añañuca" (*Zephyranthes phycelloides*), formerly known as *Rhodophiala phycelloides*. Many species of *Zephyranthes* in the Atacama Desert are dangerously threatened, due to massive extraction of bulbs and cutting of flowers. Therefore, studies of the biodiversity of these endemic species, which are essential for their conservation, should be conducted sooner rather than later. There are some chloroplast genomes available for *Amaryllidaceae* species, however there is no complete chloroplast genome available for any of the species of *Zephyranthes* subgenus *Myostemma*. The aim of the present work was to characterize and analyze the chloroplast of *Z. phycelloides* by NGS sequencing. The chloroplast genome of the *Z. phycelloides* consists of 158,107 bp, with typical quadripartite structures: a large single copy (LSC, 86,129 bp), a small single copy (SSC, 18,352 bp), and two inverted repeats (IR, 26,813 bp). One hundred thirty-seven genes were identified: 87 coding genes, 8 rRNA, 38 tRNA and 4 pseudogenes. The number of SSRs was 64 in *Z. phycelloides* and a total of 43 repeats were detected. The phylogenetic analysis of *Z. phycelloides* shows a distinct subclade with respect to *Z. mesochloa*. The average nucleotide variability (*Pi*) between *Z. phycelloides* and *Z. mesochloa* was of 0.02000, and seven loci with high variability were identified: $psbA$, $trnS^{GCU}$-$trnG^{UCC}$, $trnD^{GUC}$-$trnY^{GUA}$, $trnL^{UAA}$-$trnF^{GAA}$, $rbcL$, $psbE$-$petL$ and $ndhG$-$ndhI$. The differences between the species are furthermore confirmed by the high amount of SNPs between these two species. Here, we report for the first time the complete cp genome of one species of the *Zephyranthes* subgenus *Myostemma*, which can be used for phylogenetic and population genomic studies.

1. Introduction

Zephyranthes phycelloides or "Red añañuca" (recognizable by its red flowers) is distributed in the Atacama, Coquimbo, Metropolitan and Maule regions, and has an altitudinal range from 0 to 2200 m (Rodriguez et al., 2018). In the Atacama Desert, specifically in the Atacama Region from Chile, *Z. phycelloides* germinate, flower and reproduce in a short period of time due to a phenomenon called "Desierto Florido", which is triggered by a winter precipitation greater than 15 mm, associated with the El Niño-Southern Oscillation (ENSO) weather phenomenon (Gutiérrez, 2008). Most of the plant species that emerge during these events are endemic and exclusive to the Atacama Desert (Manrique et al., 2014, Contreras et al., 2020a), and have barely or not been studied before.

Zephyranthes phycelloides (Herb.) Nic.García, formally known as *Rhodophiala phycelloides* (Herb.) Hunz belongs to the highly polyphyletic family *Amaryllidaceae* J. St.-Hil., a group of monocotyledonous, geophytic, bulbous, petaloid, cosmopolitan plants (Meerow et al., 2000). The family is composed of 1600 species of approximately 75 genera and is widely distributed in South America, the Mediterranean and South Africa (Xu & Chang, 2017). Part of the genus *Rhodophiala* Presl. was recently re-classified as *Zephyranthes* subgenus *Myostemma* (Salisb.) Nic.García, and it has been estimated that approximately 17 species belong to this new subgenus

(García et al., 2019). The species belonging to this new subgenus grow in Chile and Argentina between 24° S and 42° S, from deserts to Patagonian steppe.

The family *Amaryllidaceae* is notoriously complicated in terms of diagnosability if the origin of the individual is unknown, as the family presents a high rate of both polyphyly and hybridization (García et al., 2014, 2017, 2019). The levels of polyphyly and hybridization of the family and the genus *Rhodophiala* Presl. have been studied with karyotyping and phylogenetic approaches (Muñoz et al., 2007, Baeza et al., 2012), but there is still no consensus in its determination (Baeza et al., 2016). Comparison of the karyotypes show that *Z. phycelloides*, *Zephyranthes bagnoldii* (Herb.) Nic.García and *Zephyranthes advena* (Ker Gawl.) Nic.García all possess a 2n = 18 karyotype and share the same morphometry (Baeza et al., 2012) and molecular studies of internal transcribes spacer (ITS) sequences place *Z. phycelloides* in a closely related monophyletic group together with *Z. bagnoldii*, *Zephyranthes montana* (Phil.) Nic.García, *Zephyranthes splendens* (Renjifo) Nic.García and *Zephyranthes ananuca* (Phil.) Nic.García (Muñoz et al., 2011). Phylogenetic studies in species of the tribe *Hippeastreae* (family *Amaryllidaceae*), suggest natural hybridizations might have occurred between species from the *Zephyranthes* subg. *Myostemma* and *Hippeastrum* genera (García et al., 2014). Genetic analysis can shed a light on the genetic diversity of these endemic species, which can be used for their conservation.

Chloroplast genomes are highly valuable in taxonomy, as they are mainly maternally inherited and highly conserved (Zhang et al., 2016; Chávez-Galarza, 2021). Their slow evolution rate compared to the nuclear genome informs about molecular evolution, RNA editing, and population genetics and can solve inter-species relationships in phylogenetic studies (Zhang et al., 2018). There are some chloroplast genomes are available for *Amaryllidaceae* species (i.e. *Zephyranthes mesochloa* Herb. ex Lindl. (Namgung et al., 2021), *Narcissus poeticus* L. (Könyves et al., 2018), *Lycoris longituba* Y.C.Hsu & G.J.Fan (F. Zhang et al., 2019), *Hippeastrum vittatum* (L'Her.) Herb. (Li et al., 2020), *Hippeastrum rutilum* (Ker Gawl.) Herb. (Huang, 2020), and the species of the subfamily *Allioideae* Herbert (Namgung et al., 2021)). However, so far, there was no complete chloroplast genome available for any of the species of

Zephyranthes subgenus *Myostemma*. We bridged this void and characterized and analyzed the chloroplast of *Z. phycelloides* with NGS sequencing. Our results can serve in further chloroplast analysis and phylogeny studies of the species from the *Zephyranthes* subgenus *Myostemma* genus, but also to disentangle the genetic and evolutionary complexities of the *Amaryllidaceae* family in future studies.

2. Materials and methods

2.1. Plant material and genomic DNA isolation

Fresh leaves of *Z. phycelloides* were collected near Totoral (a small town located at 27°55′15.53″S 70°56′33.67″W) (Map of the Atacama Region in Fig. 1; MINEDUC, 2021). DNA was isolated from the leaves using the modified cetyl-trimethylammonium bromide (CTAB) protocol (Contreras et al., 2020b). The DNA was quantified with a Qubit™ 3.0 fluorometer and a Qubit™ dsDNA HS Assay Kit (Life Technologies, San Diego, CA), according to the protocol provided by the manufacturer. DNA integrity was verified with an Agilent 2100 Bioanalyzer (Agilent Technologies, San Diego, CA) prior to sequencing.

2.2. Genome sequencing, assembling and annotation

Sequencing libraries were generated by a TruSeq Nano DNA LT Kit (Illumina, San Diego, CA). The final libraries were run on an Agilent 2100 Bioanalyzer to verify the fragment size distribution and concentration. Sequencing was performed at Genoma Mayor (Universidad Mayor, Chile) with an Illumina sequencing platform. Paired-end sequences of 150 bp were generated for each read (R1 and R2). The filtered reads were assembled using SPAdes 4 software version 3.13.0 (Bankevich et al., 2012), using three k-mers parameters: - k 33, 55 and 77. The chloroplast was annotated with DOGMA software (Wyman et al., 2004) and CPGAVAS2 (Shi et al., 2019), and then manually corrected. The graphical map of the chloroplast was generated by Organellar Genome DRAW (OGDRAW) (Greiner et al., 2019), and the complete nucleotide
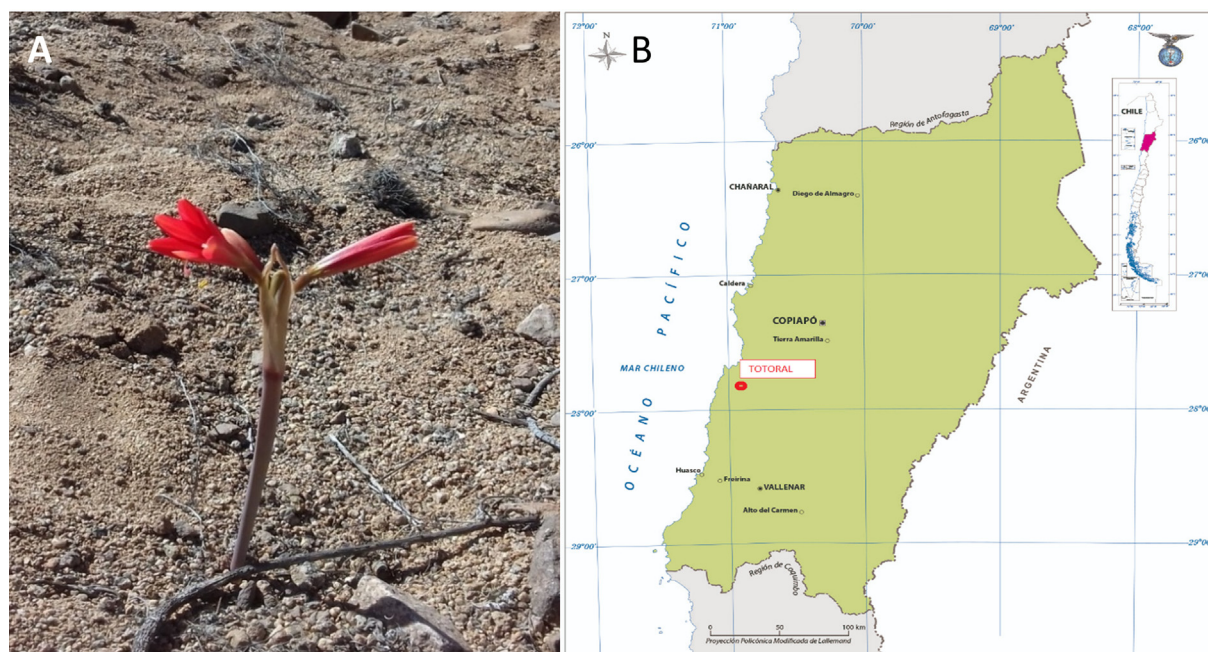

Fig. 1. Photo of *Zephyranthes phycelloides* (A) and Map of the Atacama Region (MINEDUC, 2021), showed localization of the Totoral town (B).

sequence was deposited in the GenBank database (MW348956.1, under the name *R. phycelloides*)

2.3. Genome comparison, repeat and phylogenetic analysis

The chloroplast structures (LSC/IR, IR/SSC) of *Z. phycelloides* and a total of nine closely related species in the *Amaryllidaceae* (*Hippeastrum rutilum* (Ker Gawl.) Herb., *Hippeastrum vittatum* (L'Hér.) Herb., *Narcissus poeticus* L., *Lycoris sprengeri* Comes ex Baker, *Lycoris sanguinea* Maxim., *Lycoris radiata* (L'Hér.) Herb., *Lycoris aurea* (L'Hér.) Herb., *Clivia miniata* (Lindl.) Regel and *Zephyranthes mesochloa* Herb. ex Lindl.) of the order *Asparagales* were visualized and compared using IRScope (Amiryousefi et al., 2018). As the subtribe *Hippeastrinae* is divided in two genera, *Hippeastrum* and *Zephyranthes* (García et al., 2019), we use the sequence of whole plastome (GenBank) of this subtribe for identification of simple sequence repeat (SSR). MISA software (Beier et al., 2017) was used to identify SSR in the chloroplast genome of the subtribe *Hippeastrinae*, with following search parameters: ≥10 repeat units for mononucleotide SSRs; ≥5 repeat units for dinucleotide SSRs; ≥4 repeat units for trinucleotide SSRs; and ≥ 3 repeat units each for tetra-, penta-, and hexanucleotide SSRs. To identify the tandem repeats (forward, palindromic, reverse, and complement) of these species, REPuter (Kurtz & Schleiermacher, 1999) was used. In addition, a sliding window analysis (window length: 600 pb, step size: 200 bp) was performed to assess the variability (Pi) between *Z. phycelloides* and *Z. mesochloa* chloroplasts with DnaSP v. 5 software (Librado & Rozas, 2009). The complete chloroplast genome sequence of the ten species were aligned using MAFFT v7 (Katoh & Standley, 2013). The genome sequence data were analyzed using the maximum likelihood (ML) and the Bayesian inference (BI) methods. The best-fitting nucleotide substitution model of sequence evolution, model GTR + I + G, was determined using the Akaike Information Criterion (AIC) with MrModeltest v2.3 (Nylander, 2004). The ML analyses were performed using RAxML-HPC BlackBox v.8.1.24 (Stamatakis, 2014) with 1,000 bootstrap replicates; and the BI analysis was conducted using MrBayes v.3.2 (Ronquist et al., 2012) with the CIPRES Science Gateway v3.3 (Miller et al., 2010). The Markov chain Monte Carlo (MCMC) algorithm was calculated for 5,000,000 generations, and the sampling tree for every 1,000 generations. The first 25% of generations were discarded as burn-in. In the analysis, bootstrap (BS) values were estimated in the ML, and the reliability of clades in the Bayesian analysis was evaluated by means of posterior probability (PP). The trees were visualized with FigTree (Rambaut, 2012). We additionally compared the single nucleotide polymorphism (SNP) loci of the chloroplast genome of *Z. phycelloides* with species from order *Asparagales*.
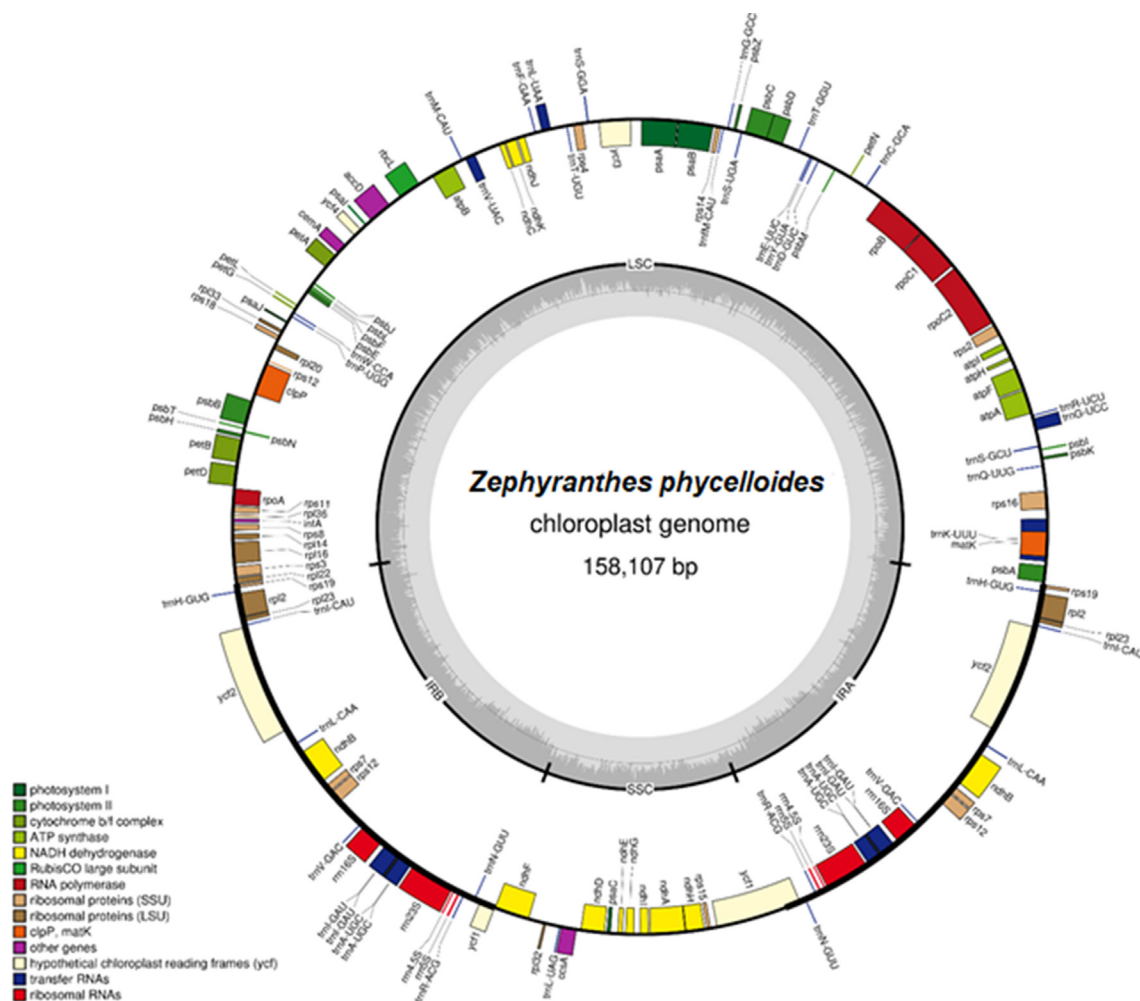

Fig. 2. Circular gene map of the chloroplast genomes of *Zephyranthes phycelloides*. Genes were colored according to their functional group. The GC content is represented by the dashed darker grey area in the inner circle, the lighter grey area represents AT content. Small single copy (SSC), large single copy (LSC), and inverted repeats (IRA, IRB) were indicated.

Table 1
General features of chloroplast genomes.

Species	Accession	Size (bp)	GC (%)	LSC (bp)	SSC (bp)	IR (bp)	No Genes
Zephyranthes phycelloides	MW348956.1	158,107	38.0	86,129	18,352	26,813	137
Zephyranthes mesochloa	MT323238.1	158,768	38.0	86,421	18,121	27,113	135
Hippeastrum rutilum	MT133568.1	158,357	37.9	86,451	18,272	26,817	133
Hippeastrum vittatum	MT762362.1	158,082	37.9	86,166	18,284	26,816	137
Narcissus poeticus	NC_039825.1	160,099	37.8	86,445	16,434	28,610	132
Lycoris sprengeri	MN158986.1	158,687	37.8	86,490	18,541	26,828	137
Lycoris sanguinea	NC_047453.1	158,761	37.7	86,529	18,430	26,901	137
Lycoris radiata	NC_045077.1	158,335	37.8	86,613	18,262	26,730	137
Lycoris aurea	NC_046752.1	158,690	37.7	86,585	18,541	26,782	132
Clivia miniata	MN857162.1	158,114	38.0	86,204	18,334	26,788	133

3. Results

The chloroplast of *Z. phycelloides* comprises 158,107 bp and its structure contains a typical quadripartite structure with two inverted repeat regions (IRs; 26,813 bp) separated by a large single copy region (LSC; 86,129 bp) and a small single copy region (SSC; 18,352 bp) (Fig. 2, Table 1). Its length and structure are similar to those of other species of the order *Asparagales*, who vary in the IRs between 26,730 bp and 28,610 bp, in the LSC between 86,129 bp and 86,613 bp and in the SSC between 18,121 bp and 18,541 bp (Table 1). *Z. phycelloides* had 25 bp more than *H. vittatum,* 250 bp less than *H. rutilum* and 661 bp less than *Z. mesochloa* (Table 1). The GC content of *Z. phycelloides* (38%) was similar to other species of the order *Asparagales* (Table 1).

The cp genome of *Z. phycelloides* contained 137 genes in total, including 87 coding genes, 8 rRNA genes, 38 tRNA genes and 4 pseudogenes (Table 2). The *ycf15* and *ycf68* are pseudogenes as they present several internal stop codons within the coding regions. Seven coding genes (*ndhB, rpl2, rpl23, rps19, rps12, rps7* and *ycf2*), four rRNA genes, eight tRNA genes, and the two above mentioned pseudogenes (*ycf15* and *ycf68*), located in the IR regions, contained duplicated genes (Table 2). Nineteen of the 137 genes contained at least one intron (Table 2).

We compared the simple sequence repeats (SSRs) from the cp genomes of the four species of the subtribe *Hippeastrinae*. The maximum number of SSR was 67 SSRs in *H. vittatum*, 64 SSRs in *Z. phycelloides* and 58 SSRs in *Z. mesochloa*, whereas *H. rutilum* had only 56 SSRs (Fig. 3A). Mononucleotide repeats were the most common repeats i.e. 46 SSRs in *H. vittatum*, 43 SSRs in *Z. phycelloides*, 36 SSRs in *Z. mesochloa* and 35 SSRs in *H. rutilum* (Fig. 3A), of which the A/T repeat was most abundant (Fig. 3B) with 45 and 41 repeats in *H. vittatum* and *Z. phycelloides*, respectively. The number of dinucleotide SSRs (AG/CT and AT/AT) was similar in all of the four species. Pentanucleotide SSRs (AAACG/CGTTT) occured in all species, except in *Z. phycelloides* (Fig. 3B), and only one hexanucleotide SSR (ACATAT/ATATGT) was observed in *Z. phycelloides*.

A total of 43, 37, 37 and 38 repeats were detected in the *Z. phycelloides*, *Z. mesochloa*, *H. rutilum* and *H. vittatum* cp genomes, respectively (Fig. 4A). The repeats in *Z. phycelloides* contained 23 palindromic, 16 forward, 3 reverse and one complement repeat, while the *Z. mesochloa* cp genome contained 22 palindromic, 15 forward, zero reverse and zero complement repeats (Fig. 4A). Palindromic and forward repeats with lengths between 30 and 40 bp were most common and abundant in the four species. However, the number of palindromic and forward repeats with lengths between 30 and 40 bp was a little higher in the cp genome of

Table 2
Gene composition in the *Zephyranthes phycelloides* chloroplast genome.

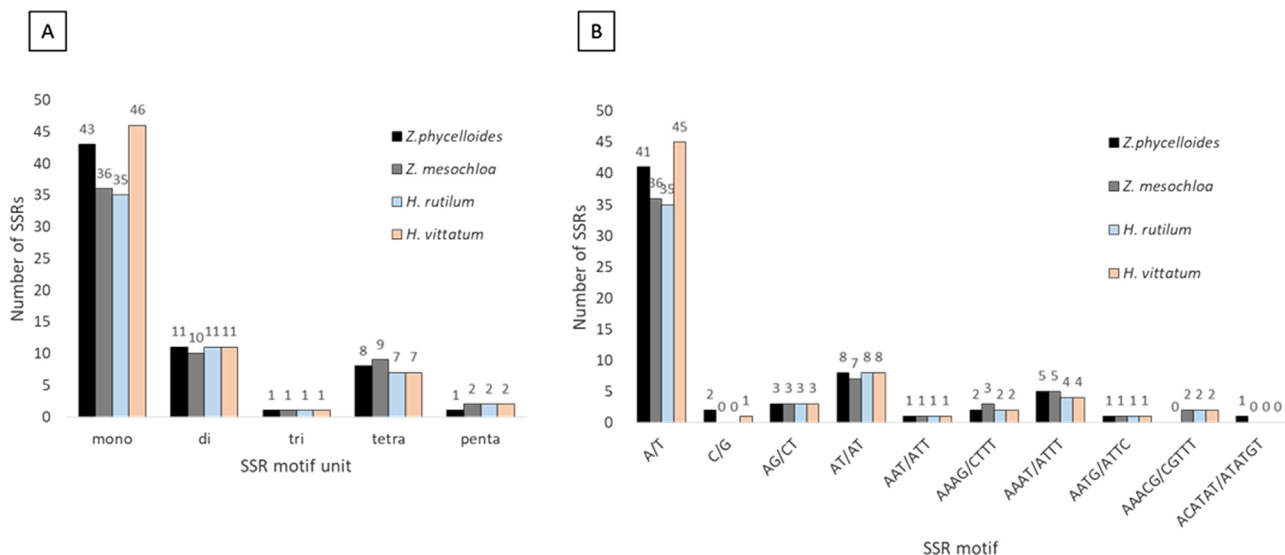| Functions	Group of genes	Name of genes	N°
Photosynthesis	Photosystem I	*psaA, psaB, psaC, psaI, psaJ*	5
	Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*	15
	ATP synthase	*atpA, atpB, atpE, atpF^a, atpH, atpI*	6
	NADH-dehydrogenase	*ndhA^a, ndhB^a(2x), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*	12
	cytochrome *b*/f complex	*petA, petB, petD^a, petG, petL, petN*	6
	Large subunit RUBISCO	*rbcL*	1
Protein synthesis and DNA replication	Transfer RNAs	*trnA-UGC^a(2x), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-UCC^a, trnG-GCC, trnH-GUG (2x), trnI-GAU^a(2x), trnI-CAU (2x), trnK-UUU^a, trnL-UAA^a, trnL-CAA (2x), trnL-UAG, trnM-CAU, trnN-GUU (2x), trnP-UGG, trnQ-UUG, trnR-ACG (2x), trnR-UCU, trnS-GGA, trnS-UGA, trnS-GCU, trnT-GGU, trnT-UGU, trnV-UAC^a, trnV-GAC (2x), trnW-CCA, trnY-GUA*	38
	Ribosomal RNAs	*rrn16S (2x), rrn23S (2x), rrn4.5S (2x), rrn5S (2x)*	8
	Ribosomal Protein large-subunit	*rpl14, rpl16, rpl2^a(2x), rpl20, rpl22, rpl23 (2x), rpl32, rpl33, rpl36*	11
	DNA dependent RNA polymerase	*rpoA, rpoB, rpoC1^a, rpoC2*	4
	Ribosomal Protein Small-subunit	*rps11, rps12^a (2x), rps14, rps15, rps16, rps18, rps19 (2x), rps2, rps3, rps4, rps7 (2x), rps8*	15
Other functions	Subunit of Acetyl-CoA-carboxylase	*accD*	6
	c-type cytochrom synthesis gene	*ccsA*	
	Envelop membrane protein	*cemA*	
	Protease	*clpP*	
	Maturase	*matK*	
	Initiation Factor	*infA*	
Unknown function	Conserved open reading frames	*ycf1 (2x), ycf2 (2x), ycf3^a, ycf4, ycf15 (2x), ycf68 (2x)*	10

(2x) Duplicated genes; (a) Genes containing introns.

Fig. 3. Analysis of simple sequence repeats (SSRs) of the *Z. phycelloides, Z. mesochloa, H. rutilum* and *H. vittatum* chloroplast genomes. Total numbers of SSRs of each motif unit (A) and number of SSRs detected of each motif type (B).
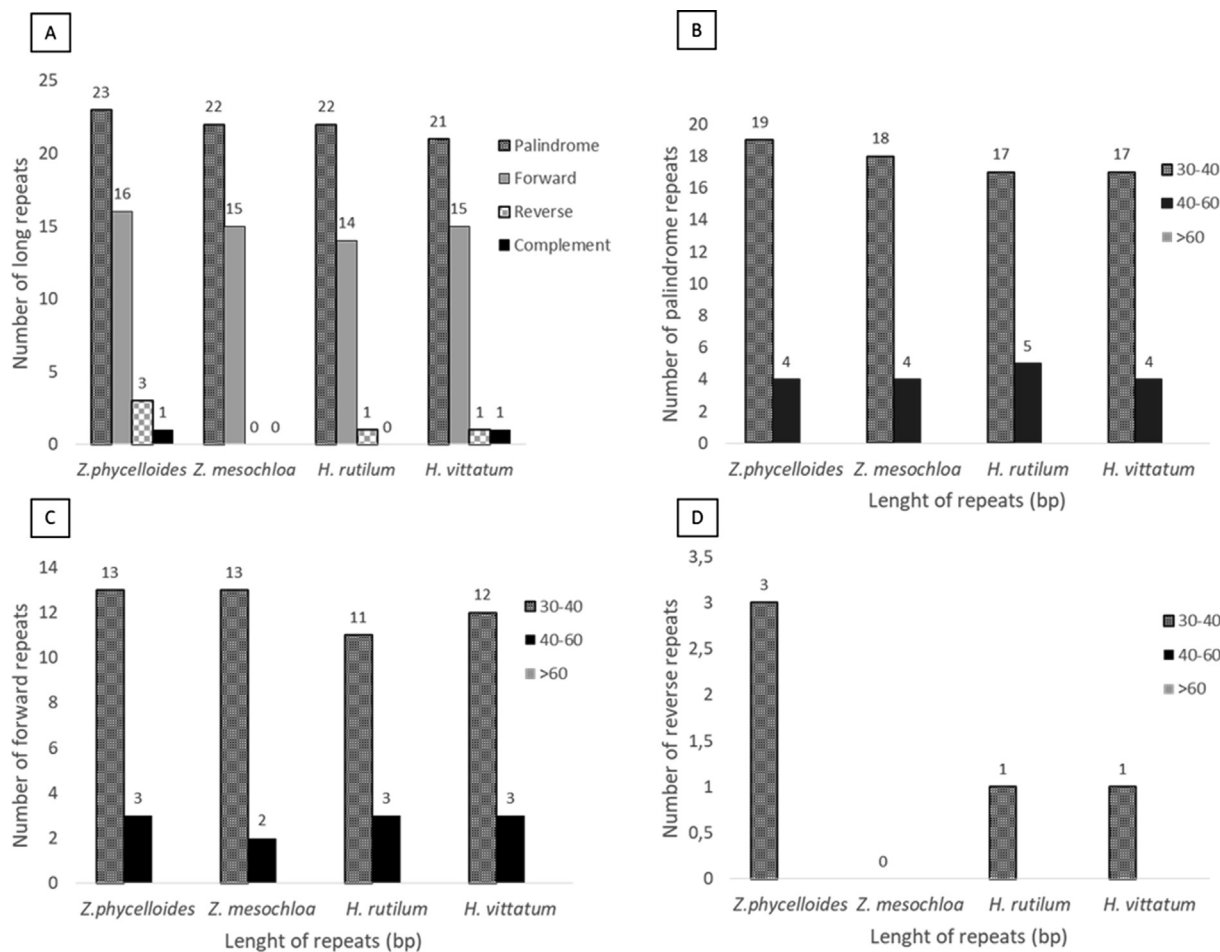

Fig. 4. Repeat structure analysis of the *Z. phycelloides, Z. mesochloa, H. rutilum* and *H. vittatum* chloroplast genomes. Total numbers long repeat types (Palindrome, Forward, Reverse and Complement) (A), number of palindrome repeats (B), number of forward repeats (C) and number of reverse repeats (D) by length.

Zephyranthes than in the *Hippeastrum* genus. The number of palindromic and forward repeats with length between 40 and 60 bp was similar in both genera (Fig. 4BC). The reverse repeat (range of 30–40 bp) was more abundant in *Z. phycelloides* than in the rest of the species (Fig. 4D).

Although normally IR regions have similar lengths within the chloroplast (Gruenstaeudl & Jenke, 2020) it has been shown that IR regions can expand or contract (Ogihara et al., 2001, Odintsova & Yurina, 2007). For that reason, we compared the information of the IR-SSC and IR-LSC limits of *Z. phycelloides* with other species
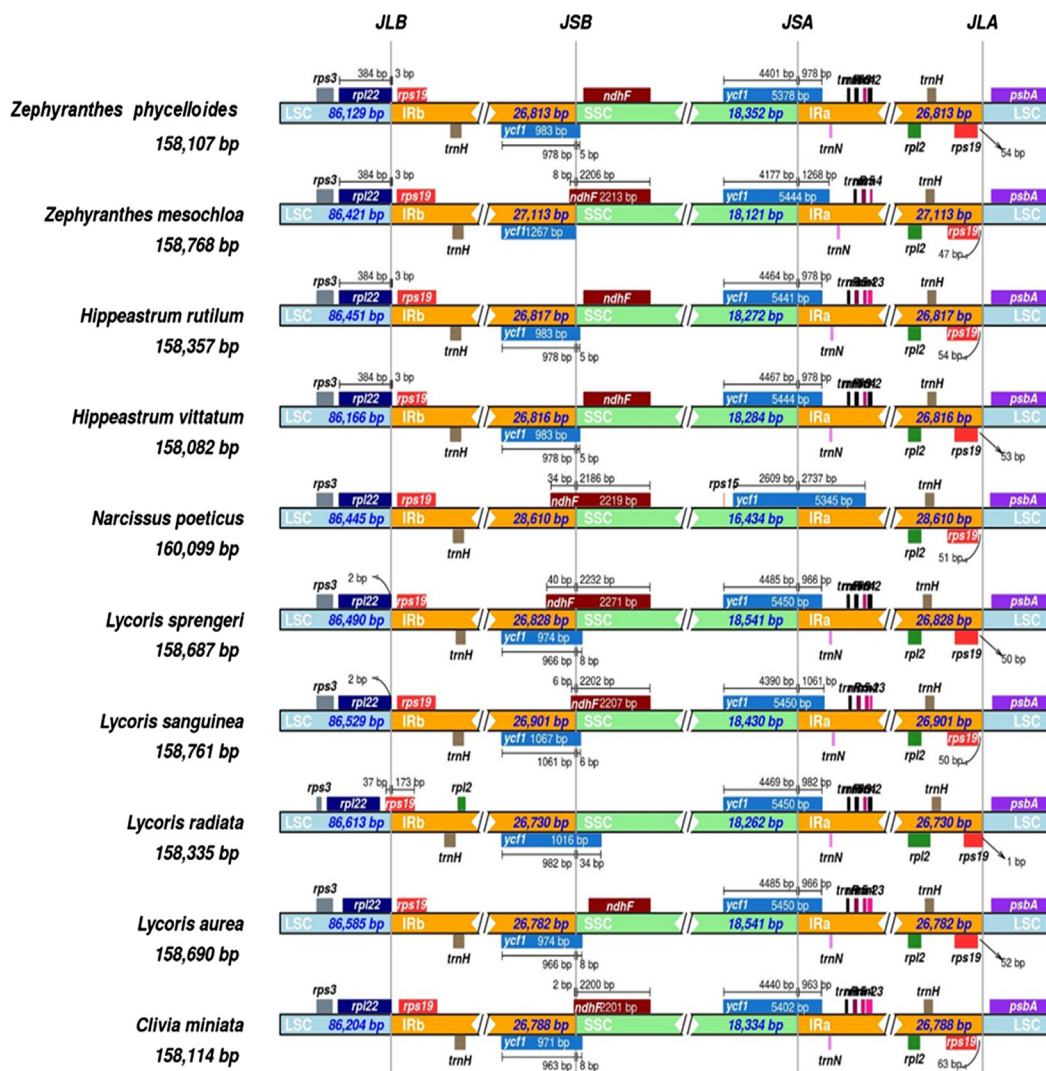
Fig. 5. Comparison of chloroplast genomes between the Long Single Copy (LSC), Short Single Copy (SSC) and Inverted Repeat (IRa and IRb) junction regions amongst ten species of the order *Asparagales*.
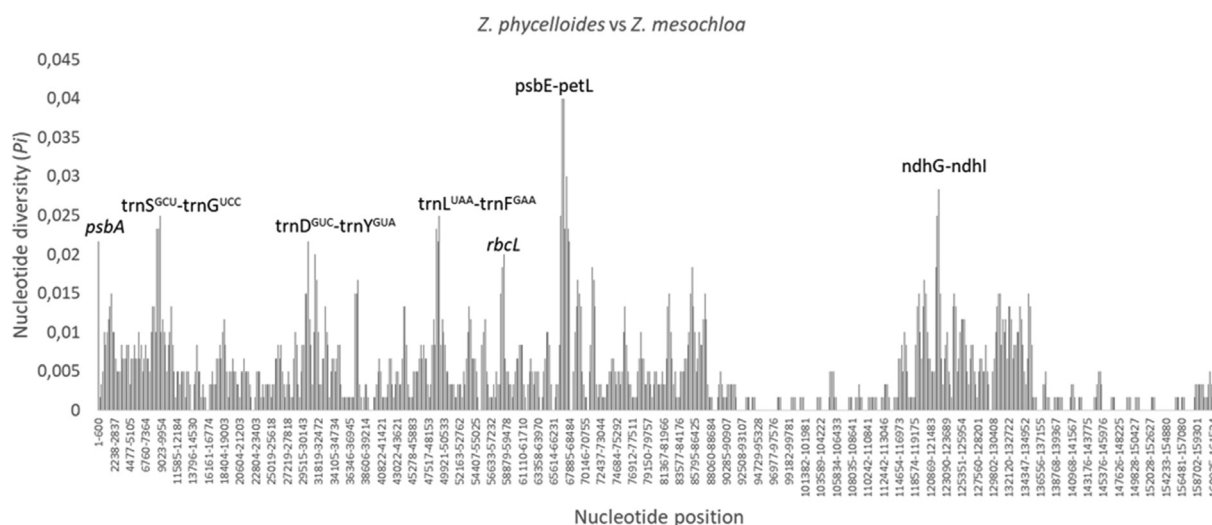

Fig. 6. Sliding window analysis of the whole chloroplast of *Z. phycelloides* and *Z. mesochloa*. (Window length:600 bp, step size: 200 bp). X-axis: Nucleotide position, Y-axis: Nucleotide diversity (Pi).

from the order *Asparagales*. The intergenic spacers (IGS) between *rpl22-rps19* genes of *Z. phycelloides*, in the junction between LSC and IRb region (JLB), are similar in size to those of *Hippeastrum* species and *Z. mesochloa*, while there is more variation when compared with other *Asparagales* species (Fig. 5). Likewise, the intergenic spacers (IGS) between *rps19-psbA* genes, located in the JLA junction, of *Z. phycelloides* and *Hippeastrum* species are similar in size, but they differ from *Z. mesochloa* and the other *Asparagales* species (Fig. 5). The boundaries between IRa and SSC (JSA) were located in the *ycf1* gene, and the fragment located in the IRa region was equal in size compared to *Hippeastrum* species (978 bp), whereas it differed from *Z. mesochloa* (1268 bp) and the other

Asparagales species (940 bp to 2737 bp) (Fig. 5). Similarly, the *ycf1* gene spanned the IRb/SSC region, and the fragment located at the IRb region was equal in size compared to *Hippeastrum* species (978 bp), whereas it varied between *Z. mesochloa* (1267 bp) and the other *Asparagales* species (between 963 bp and 1061 bp) (Fig. 5). Additionally, when comparing the four *Hippeastrinae* species, only *Z. mesochloa* showed differences from the others in the JSB junction (*ndhF* gene) (Fig. 5).

The average nucleotide variability (*Pi*) between *Z. phycelloides* and *Z. mesochloa* was estimated to be 0,00492 (ranging from 0 to 0.04000) (Fig. 6). The most variable regions in the chloroplast were located in LSC and SSC regions, whereas the IR regions had a much
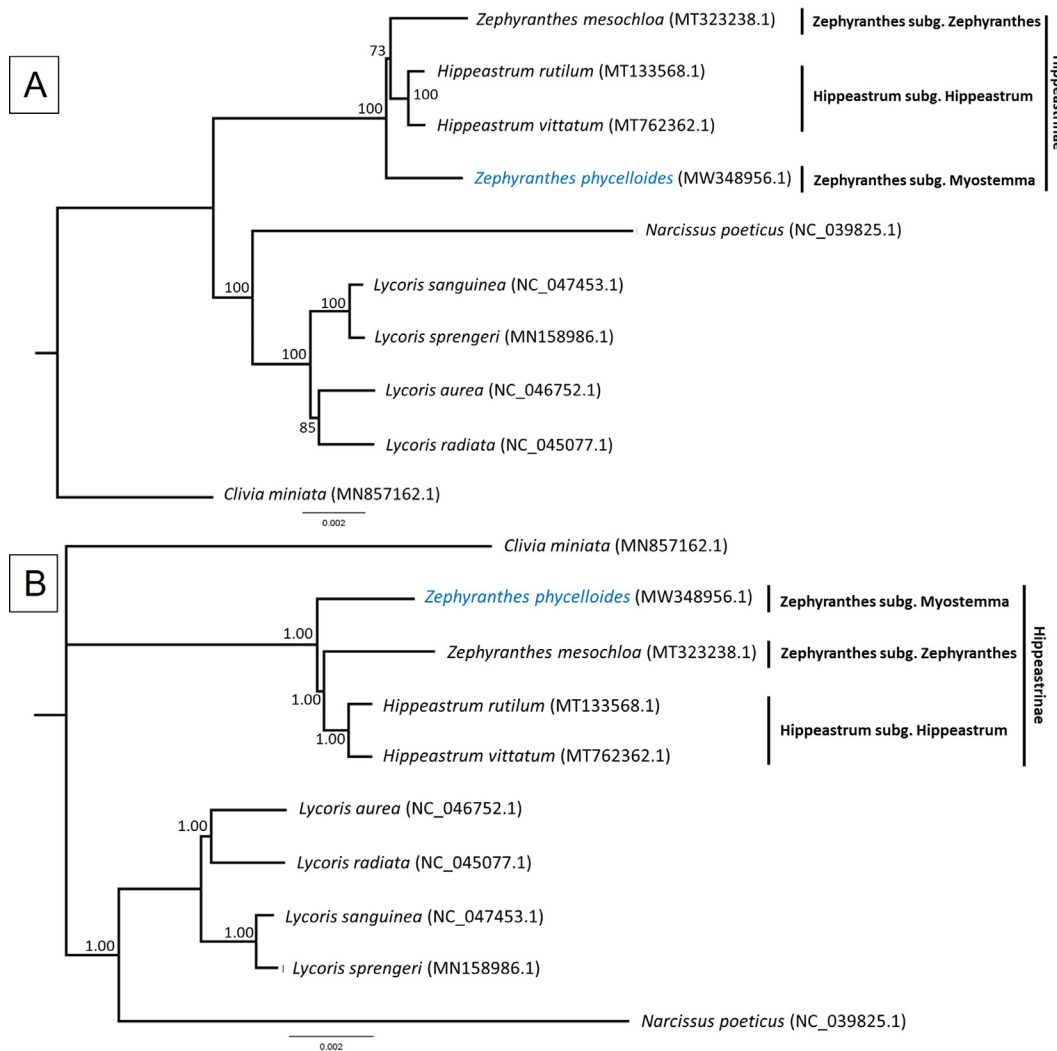

Fig. 7. Molecular phylogenetic analysis of ten whole chloroplast genomes inferred by ML (A) and BI (B) methods. Numbers in the branches are ML bootstrap values (BS) on the above tree (A) and Bayesian posterior probabilities values (PP) on the below tree (B).

Table 3
Number of SNPs for six chloroplast genome.

	Z. phycelloides	Z. mesochloa	H. rutilum	H. vittatum	N. poeticus	L. aurea
Z. phycelloides	–					
Z. mesochloa	776	–				
H. rutilum	549	553	–			
H. vittatum	558	559	172	–		
N. poeticus	2901	2966	2774	2763	–	
L. aurea	1876	1877	1674	1669	2212	–

lower nucleotide diversity. Seven loci with high levels of variability (Pi greater than 0.02000) were found: *psbA* (Pi = 0.02167), *trnS^{GCU}-trnG^{UCC}* (Pi = 0.02333), *trnD^{GUC}-trnY^{GUA}* (Pi = 0.02000), *trnL^{UAA}-trnF^{GAA}* (Pi = 0.02333), *rbcL* (Pi = 0, 02000), *psbE-petL* (Pi = 0.0400) and *ndhG-ndhI* (Pi = 0.02833) (Fig. 6).

The results of the ML and BI trees had similar topologies when we compared the whole chloroplast genome sequences of the ten species of the order *Asparagales* (Fig. 7). The ML phylogenetic analysis revealed four clades, one joined *Z. phycelloides, Z. mesochloa, H. rutilum* and *H. vittatum* (BP = 100), the second clade contained *N. poeticus* (BP = 100), the third clade was formed by *Lycoris* species (BP = 100) and the fourth clade was formed by the outgroup *C. miniata* (Fig. 7A). The BI phylogenetic analysis from the species that form the clade *Hippeastrinae* showed high support (PP = 1.00) and their topology was identical to ML analysis (Fig. 7B). The first clade showed two subclades where the two species of the genus *Hippeastrum* (*H. vittatum* and *H. rutilum*) were joined to *Z. mesochloa* with strong support (ML, BP = 73; BI, PP = 1.00), while *Z. phycelloides* (*Zephyranthes* subg *Myostemma*) was separated from them with high support (BP = 100; PP = 1.00) (Fig. 7). *Z. phycelloides* differed in 776, 549, 558, 2901 and 1876 SNPs from the chloroplasts of *Z. mesochloa, H. rutilum, H. vittatum, N. poeticus* and *L. aurea* respectively, while the chloroplast genome substitution events between *H. rutilum* and *H. vittatum* were only 172 SNPs (Table 3).

4. Discussion

We found that the complete chloroplast genome of *Z. phycelloides*, is highly conserved in size, number of genes and percentage of GC content in comparison to other species of the order *Asparagales*. Our results are consistent with other studies, which found that the chloroplast genome and the GC content, gene composition and order of genes of the *Amaryllidaceae* family are highly conserved (Jimenez et al., 2020, Namgung et al., 2021). We found that *Z. phycelloides*, and other species of the subTribe *Hippeastrinae* (subfamily *Amaryllidoideae*) that were used in this study, presented more GC content (~38%) than the species of the subfamily *Allioideae* (≤37.1%; Namgung et al., 2021). Interestingly, the genome size of *Z. phycelloides* differed more from *Z. mesochloa* than from *H. vitattum* and *H. rutilum*. This difference between *Z. mesochloa* and *Z. phycelloides* is due to expansion and contraction of the *ycf1* and *ndhF* genes in the IRs regions, confirming that IRs are highly variable due to lineage-specific expansions and contractions (Zhu et al., 2016).

Zephyranthes phycelloides possesses a similar amount of genes as *H. vittatum,* whereas *Z. mesochloa* has 2 genes less (MT323238.1; Namgung et al., 2021). This variation in the number of genes could be explained by pseudogenization and gene loss that can occur in species (Petersen et al., 2015, Li et al., 2017, Mohanta et al., 2020). For example, *Z. phycelloides* possesses two copies of the pseudogene *ycf68* that are not observed in the annotation of *Z. mesochloa* (Namgung et al., 2021). Pseudogenization of *rps2* and gene loss of *infA* causes chloroplast genome fluctuations among *Allioideae* species, whereas pseudogenization of *ycf15* causes the fluctuations in *Amaryllidaceae* species (Namgung et al., 2021). However, to confirm if pseudogenization and gene loss are common in species of the genus *Zephyranthes* it will be necessary to analyze more chloroplasts.

A total of 56 to 67 chloroplast simple sequence repeats (cpSSRs) were founded in the cp genomes of *Z. phycelloides, Z. mesochloa, H. vittatum* and *H. rutilum* (*Hippeastrinae* species). Our results showed variation in the number of cpSSRs, and minor differences in the mono-di-tri-tetra-penta and hexa motifs. Chloroplast simple sequence repeats (cpSSRs) are widely used to study phylogeny, but they can also be used in ecological studies or in evolutionary

processes (Daniell et al., 2016). For example, 72 cpSSRs (57 were mononucleotide, 13 dinucleotide and 2 trinucleotide) were discovered in 17 cp genomes of *Allioideae* (*Amaryllidaceae*) species, using similar SSRs search parameters as in our study (Namgung et al., 2021). A number of 44 to 54 cpSSRs were found in seven *Lycoris* species (*Amaryllidaceae*), however the SSRs search parameters were more stringent in this study (Zhang et al., 2020). Consequently, the unique genomic coding of cpSSRs of a species can be used to analyze the genetic variation of populations and evolutionary processes of their genus.

We found the highest number of repeats elements in *Z. phycelloides* (43) and the lowest in *H. vittatum* (38). These repeat elements that are important for promoting genetic rearrangements (Wu et al., 2017). Within the *Hippeastrinae* species the palindromic repeat was the most common one. Similar as in our species, a total between 37 and 49 repeats were found in nine cp genomes of the *Allium* genus, however the number of palindromic type repeats was lower (Huo et al., 2019). Another study found 21 repeats in 17 cp genomes of *Allioideae* species (Namgung et al., 2021). These repeats were mainly forward repeats, whereas palindromic repeats were only found in two species (Namgung et al., 2021). The high total number of repeats as well as the high number of reverse treats found in *Z. phycelloides*, are important as they give insight into genome variations, rearrangements or structural expansions within and between species (Wicke et al., 2011).

The expansion and contraction of IR regions in our species are the result of modifications in the junctions of the LSC-IR-SSC regions. These expansions and contractions cause length variations in species of the order *Asparagales* (Wang et al., 2008), especially in the subtribe *Hippeastrinae*, where these regions and their connections show clear differences between *Z. phycelloides* and *Z. mesochloa*. We found a larger expansion in the *ycf* and *ndhF* genes in the junction regions in *Z. mesochloa* than *Z. phycelloides*, making the size of the chloroplast of *Z mesochloa* larger than the chloroplast of *Z. phycelloides*. However, the junctions of *Z. phycelloides* and the two *Hippeastrum* species are very similar. I.e. in the SSC-IR border (JSB) of *Z. mesochloa* the *ndhF* and *ycf1* genes overlapped, whereas these two genes were separated by a gap in *Z. phycelloides, H. rutilum* and *H. vittatum*. This corresponds to previous studies in *Allioideae* species that found that the SSC-IR border (JSB) can be variable, showing either overlap or present a gap between *ycf1* and *ndhF* (Do et al., 2020, Namgung et al., 2021). We did not find the *rps19* gene within the IR-LSC borders (JLB and JLA) in any of the four *Hippeastrinae* species. However, Zhang et al. (2020), did detect the *rps19* gene within the JLB border in two *Lycoris* species (*Amaryllidaceae*), showing a partial duplication, while five other species showed an IR expansion, and thus a complete duplication of *rps19*. Successive IR expansions have shown the importance of the JLA and JLB junctions for the analysis of evolutionary processes, providing clues about the origin and evolution of species (Palmer and Stein, 1986, Goulding et al., 1996).

In the present work, seven regions with high level of variability (Pi greater than 0.02) were identified: *psbA, trnS^{GCU}-trnG^{UCC}, trnD^{GUC}-trnY^{GUA}, trnL^{UAA}-trnF^{GAA}, rbcL, psbE-petL* and *ndhG-ndhI*. Chloroplast markers 3'*ycf1, ndhF, trnL^{UAA}-F^{GAA}* and the nuclear marker ITS rDNA have proved useful in resolving part of the phylogeny of the different genera of the tribe *Hippeastreae* (García et al., 2014). A subsequent study uncovered the information of 18 nuclear loci and 40 nearly complete chloroplast genomes for the same purpose (García et al., 2017). Even though this helped to clarify some of the complexities of the family *Amaryllidaceae*, especially in the subtribe *Hippeastrinae*, the phylogeny of the genus *Zephyranthes* was still unclear (García et al., 2019). We believe that the seven regions with high levels of variability found in this study, can be of use to resolve these uncertainties and will be useful in phylogenetic analysis of the species of the genus *Zephyranthes*.

The phylogenetic results ML and BI showed that *Z. phycelloides* represent a distinct subclade to *Z. mesochloa*, even though both belong to the same genus.. A phylogenetic analysis of near-complete chloroplasts showed a similar separation of *Z. mesochloa* from *Zephyranthes advena* (Ker Gawl.) Nic. Garcia and *Zephyranthes ananuca* (Phil.) Nic.Garcia, two other species from *Z.* subg. *Myostemma* (García et al., 2017), however, *Z. phycelloides* was not included in this work, and the more than 40 nearly complete chloroplasts are unfortunately not available for analysis. Several studies divide the *Hippeastreae* tribe into clades, based on chromosome number: *Z. phycelloides* 2n = 18, the genus *Hippeastrum* 2n = 22, and *Z. mesochloa* 2n = 12 (Greizerstein and Naranjo, 1987, Muñoz et al., 2011, García et al., 2017). Our results confirm the differences between *Z. phycelloides* and *Z. mesochloa*, which are additionally supported by SNP differences (776 SNPs). In comparison *H. vittatum* and *H. rutilum*, showed a much lower value of SNPs (172). *Allium* species (*Amaryllidaceae*) from Central Asian species contained 451 SNPs in protein-coding genes (Yusupov et al., 2020).

5. Conclusion

In this study, we report the complete sequence and characterization of the chloroplast genome of *Zephyranthes phycelloides*. Size and number of genes is conserved, similar to species from the subtribe *Hippeastrinae*. However, *Z. phycelloides* and *Z. mesochloa* showed differences in genome size and slight differences in gene number. The phylogenetic analysis showed that *Z. phycelloides* represent a distinct subclade than *Z. mesochloa*. The differences between the species are furthermore confirmed by the high amount of SNPs between these two species. The information of the complete chloroplast genome of *Z. phycelloides* shows that phylogeny of the genus *Zephyranthes* is still uncertain, and urgently need taxonomic studies of all the species of the genus. The results of cpSSRs and repetitive sequences can be helpful for population analysis and evolutionary studies of the genus *Zephyranthes*. In addition, seven highly variable regions were detected that can be used to develop useful markers for phylogenetic analysis and to distinguish between *Zephyranthes* species.

Declaration of Competing Interest

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

References

Amiryousefi, A., Hyvönen, J., & Poczai, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics (Oxford, England)*, 34(17), 3030–3031. https://doi.org/10.1093/bioinformatics/bty220.

Baeza, C., Almendras, F., Ruíz, E., Peñailillo, P., 2012. Estudio comparativo del cariotipo en especies de Miltinea Ravenna, Phycella Lindl. y Rhodophiala C Presl (Amaryllidaceae) de Chile. *Revista de La Facultad de Ciencias Agrarias* 44 (2), 193–205.

Baeza, C.M., Peñailillo, P., Novoa, P., Rosas, M., Finot, V.L., Ruiz, E., 2016. Recuentos cromosómicos en plantas que crecen en Chile IV. *Gayana - Botanica* 73 (2), 183–190. https://doi.org/10.4067/S0717-66432016000200183.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19 (5), 455–477. https://doi.org/10.1089/cmb.2012.0021.

Beier, S., Thiel, T., Münch, T., Scholz, U., Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. Bioinformatics, 33(16), 2583–2585. https://doi.org/10.1093/bioinformatics/btx198.

Chávez-Galarza, Julio C. et al., 2021. Mitochondrial DNA Part B 6 (9), 2562–2564. https://doi.org/10.1080/23802359.2021.1944364.

Contreras, R., Aguayo, F., Porcile, V., 2020a. Phylogenetic relationships of plant species from. Bol. Latinoam. Caribe Plant. Med. Aromat. 19 (3), 300–313.

Contreras, R., van den Brink, L., Burgos, B., González, M., Gacitúa, S., 2020b. Genetic characterization of an endangered chilean endemic species, prosopis burkartii muñoz, reveals its hybrids parentage. Plants 9 (6), 1–19. https://doi.org/10.3390/plants9060744.

Daniell, H., Lin, C.S., Yu, M., Chang, W.J., 2016. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. Genome Biol. 17 (1). https://doi.org/10.1186/s13059-016-1004-2.

Do, H.D.K., Kim, C., Chase, M.W., Kim, J., 2020. Implications of plastome evolution in the true lilies (monocot order Liliales). Mol. Phylogenet. Evol. 148, 106818. https://doi.org/10.1016/j.ympev.2020.106818.

García, N., Folk, R.A., Meerow, A.W., Chamala, S., Gitzendanner, M.A., de Oliveira, R. S., Soltis, D.E., Soltis, P.S., 2017. Deep reticulation and incomplete lineage sorting obscure the diploid phylogeny of rain-lilies and allies (Amaryllidaceae tribe Hippeastreae). Mol. Phylogenet. Evol. 111, 231–247. https://doi.org/10.1016/j.ympev.2017.04.003.

García, N., Meerow, A.W., Arroyo-Leuenberger, S., Oliveira, R.S., Dutilh, J., Soltis, P., Judd, W., 2019. Generic classification of Amaryllidaceae tribe Hippeastreae. Taxon 68 (3), 481–498. https://doi.org/10.1002/tax.v68.310.1002/tax.12062.

García, N., Meerow, A.W., Soltis, D.E., Soltis, P.S., 2014. Testing deep reticulate evolution in amaryllidaceae tribe hippeastreae (asparagales) with its and chloroplast sequence data. Syst. Bot. 39 (1), 75–89. https://doi.org/10.1600/036364414X678099.

Goulding, S.E., Olmstead, R.G., Morden, C.W., Wolfe, K.H., 1996. Ebb and flow of the chloroplast inverted repeat. Mol. Gen. Genet. 252 (1–2), 195–206. https://doi.org/10.1007/BF02173220.

Greiner, S., Lehwark, P., Bock, R., 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 47 (W1), W59–W64. https://doi.org/10.1093/nar/gkz238.

Greizerstein, E., Naranjo, C., 1987. Estudios cromosomicos en especies de Zephyranthes (Amaryllidacea). Darwiniana 28 (1–4), 169–186.

Gruenstaeudl, M., Jenke, N., 2020. PACVr: Plastome assembly coverage visualization in R. BMC Bioinf. 21 (1). https://doi.org/10.1186/s12859-020-3475-0.

Gutiérrez, J., 2008. El desierto Florido en la Región de Atacama. In: Squeo, F., Arancio, G., Gutiérrez, J. (Eds.), Libro rojo de la flora nativa y de los sitios prioritarios parta su conservación: Región de Atacama. Universidad de La Serena, pp. 285–291.

Huang, B., 2020. The complete chloroplast genome sequence of Hippeastrum rutilum (Amaryllidoideae). Mitochondrial DNA Part B: Resources 5 (3), 3405–3406. https://doi.org/10.1080/23802359.2020.1820388.

Huo, Y.M., Gao, L.M., Liu, B.J., Yang, Y.Y., Kong, S.P., Sun, Y.Q., Yang, Y.H., Wu, X., 2019. Complete chloroplast genome sequences of four Allium species: comparative and phylogenetic analyses. Sci. Rep. 9 (1). https://doi.org/10.1038/s41598-019-48708-x.

Jimenez, H.J., Francisco da Silva, A.D., Semen Martins, L.S., De Carvalho, R., De Moraes Filho, R.M., 2020. Comparative genomics plastomes of the Amaryllidaceae family species. Scientia Plena 16 (6).

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780. https://doi.org/10.1093/molbev/mst010.

Könyves, K., Bilsborrow, J., David, J., Culham, A., 2018. The complete chloroplast genome of Narcissus poeticus L. (Amaryllidaceae: Amaryllidoideae). *Mitochondrial DNA Part B. Resources* 3 (2), 1137–1138. https://doi.org/10.1080/23802359.2018.1521311.

Kurtz, S., Schleiermacher, C., 1999. REPuter: Fast computation of maximal repeats in complete genomes. Bioinformatics 15 (5), 426–427. https://doi.org/10.1093/bioinformatics/15.5.426.

Li, P., Ren, M., Zhu, Q., Zhang, Y., Xu, H., Wang, Z., Liu, S., Cheng, Q., Liang, B., 2020. The complete chloroplast genome of Hippeastrum vittatum (Amaryllidaceae). Mitochondrial DNA Part B: Resources 5 (3), 3539–3541. https://doi.org/10.1080/23802359.2020.1827059.

Li, Y., Zhou, J.G., Chen, X.L., Cui, Y.X., Xu, Z.C., Li, Y.H., Song, J.Y., Duan, B.Z., Yao, H., 2017. Gene losses and partial deletion of small single-copy regions of the chloroplast genomes of two hemiparasitic Taxillus species. Sci. Rep. 7 (1). https://doi.org/10.1038/s41598-017-13401-4.

Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25 (11), 1451–1452. https://doi.org/10.1093/bioinformatics/btp187.

Manrique, R., Ricotta, C., Ferrari, C., Pezzi, G., 2014. Latitudinal pattern in plant composition along the Peruvian and Chilean fog oases. Plant Biosyst. 148 (5), 1002–1008. https://doi.org/10.1080/11263504.2014.918059.

Meerow, A.W., Guy, C.L., Li, Q.B., Yang, S.L., 2000. Phylogeny of the American Amaryllidaceae based on nrDNA ITS sequences. Syst. Bot. 25 (4), 708–726. https://doi.org/10.2307/2666729.

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop, GCE 2010.* https://doi.org/10.1109/GCE.2010.5676129

MINEDUC. (2021). *Map of the Atacama Region.* Ministerio de Educación de Chile.

Mohanta, T.K., Mishra, A.K., Khan, A., Hashem, A., Abd-Allah, E.F., Al-Harrasi, A., 2020. Gene loss and evolution of the plastome. Genes 11 (10), 1–23. https://doi.org/10.3390/genes11101133.

Muñoz, M., Riegel, R., Seemann, P., Schiappacasse, F., Peñailillo, P., Jara, G., Basoalto, A., 2007. Análisis de polimorfismos intra e interespecíficos a nivel de región ITS, estudios de cariotipo y aproximaciones filogenéticas entre especies de Rhodophiala Presl. Agro Sur. 35 (2), 22–24. https://doi.org/10.4206/agrosur.2007.v35n2-11.

Muñoz, M., Riegel, R., Seemann, P., Peñailillo, P., Schiappacasse, F., Núñez, J., 2011. Phylogenetic relationships of rhodolirium montanum phil. and related species based on nucleotide sequences from ITS region and karyotype analysis. Gayana - Botanica 68 (1), 40–48. https://doi.org/10.4067/S0717-66432011000100005.

Namgung, J., Do, H.D.K., Kim, C., Choi, H.J., Kim, J.-H., 2021. Complete chloroplast genomes shed light on phylogenetic relationships, divergence time, and biogeography of Allioideae (Amaryllidaceae). Sci. Rep. 11 (1). https://doi.org/10.1038/s41598-021-82692-5.

Nylander, J. A. A. (2004). *MrModeltest Version 2.* Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Odintsova, M. S., & Yurina, N. P. (2007). Chloroplast Genomics of Land Plants and Algae. In *Biotechnological Applications of Photosynthetic Proteins: Biochips, Biosensors and Biodevices* (pp. 57–72). https://doi.org/10.1007/978-0-387-36672-2_6

Ogihara, Y., Isono, K., Kojima, T., Endo, A., Hanaoka, M., Shiina, T., Terachi, T., Utsugi, S., Murata, M., Mori, N., Takumi, S., Ikeo, K., Gojobori, T., Murai, R., Murai, K., Matsuoka, Y., Ohnishi, Y., Tajiri, H., Tsunewaki, K., 2001. Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. Mol. Genet. Genomics 266 (5), 740–746. https://doi.org/10.1007/s00438-001-0606-9.

Palmer, J.D., Stein, D.B., 1986. Conservation of chloroplast genome structure among vascular plants. Curr. Genet. 10 (11), 823–833. https://doi.org/10.1007/BF00418529.

Petersen, G., Cuenca, A., Seberg, O., 2015. Plastome evolution in hemiparasitic mistletoes. Genome Biol. Evolution 7 (9), 2520–2532. https://doi.org/10.1093/gbe/evv165.

Rambaut, A., 2012. FigTree: Tree Figure Drawing Tool Version 1.4. University of Edinburgh, Edinburgh, Scotland.

Rodriguez, R., Marticorena, C., Alarcón, D., Baeza, C., Cavieres, L., Finot, V.L., Fuentes, N., Kiessling, A., Mihoc, M., Pauchard, A., Ruiz, E., Sanchez, P., Marticorena, A., 2018. Catalogue of the vascular plants of Chile. Gayana - Botanica 75 (1), 1–430. https://doi.org/10.4067/S0717-66432018000100001.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61 (3), 539–542. https://doi.org/10.1093/sysbio/sys029.

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., Liu, C., 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 47 (W1), W65–W73. https://doi.org/10.1093/nar/gkz345.

Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30 (9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Wang, R.J., Cheng, C.L., Chang, C.C., Wu, C.L., Su, T.M., Chaw, S.M., 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol. Biol. 8 (1). https://doi.org/10.1186/1471-2148-8-36.

Wicke, S., Schneeweiss, G.M., dePamphilis, C.W., Müller, K.F., Quandt, D., 2011. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. Plant Mol. Biol. 76 (3–5), 273–297. https://doi.org/10.1007/s11103-011-9762-4.

Wu, M., Li, Q., Hu, Z., Li, X., Chen, S., 2017. The complete amomum kravanh chloroplast genome sequence and phylogenetic analysis of the commelinids. Molecules 22 (11), 1875. https://doi.org/10.3390/molecules22111875.

Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20 (17), 3252–3255. https://doi.org/10.1093/bioinformatics/bth352.

Xu, Z., & Chang, L. (2017). Identification and Control of Common Weeds: Volume 3. In *Identification and Control of Common Weeds: Volume 3* (Vol. 3). https://doi.org/10.1007/978-981-10-5403-7.

Yusupov, Z., Deng, T., Volis, S., Khassanov, F., Makhmudjanov, D., Tojibaev, K., Sun, H., 2020. Phylogenomics of Allium section Cepa (Amaryllidaceae) provides new insights on domestication of onion. Plant Diversity 43 (2), 102–110. https://doi.org/10.1016/j.pld.2020.07.008.

Zhang, F., Tong, H., Yang, H., Wang, T., Zhuang, W., Shu, X., Wang, Z., 2019. Characterisation of the complete chloroplast genome of Lycoris longituba (Amaryllidaceae). Mitochondrial DNA Part B 4 (2), 3782–3783. https://doi.org/10.1080/23802359.2019.1681324.

Zhang, F., Wang, T., Shu, X., Wang, N., Zhuang, W., Wang, Z., 2020. Complete chloroplast genomes and comparative analyses of l. Chinensis, l. anhuiensis, and l. aurea (amaryllidaceae). Int. J. Mol. Sci. 21 (16), 1–14. https://doi.org/10.3390/ijms21165729.

Zhang, X., Rong, C., Qin, L., Mo, C., Fan, L., Yan, J., Zhang, M., 2018. Complete chloroplast genome sequence of Malus hupehensis: Genome structure, comparative analysis, and phylogenetic relationships. Molecules 23 (11). https://doi.org/10.3390/molecules23112917.

Zhang, Y., Du, L., Liu, A., Chen, J., Wu, L., Hu, W., Zhang, W., Kim, K., Lee, S.C., Yang, T. J., Wang, Y., 2016. The complete chloroplast genome sequences of five Epimedium species: Lights into phylogenetic and taxonomic analyses. Front. Plant Sci. 7 (MAR2016). https://doi.org/10.3389/fpls.2016.00306.

Zhu, A., Guo, W., Gupta, S., Fan, W., Mower, J.P., 2016. Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. New Phytol. 209 (4), 1747–1756. https://doi.org/10.1111/nph.13743.