

RESEARCH ARTICLE

Complete chloroplast genomes and comparative analyses of *Hippeastrum* 'milady', *Hippeastrum albertii* and *Hippeastrum reticulatum* (Amaryllidaceae)

Xiao-fei Liu^{1,2}, Ying-bo Sun^{1,2}, Gen-fa Zhu^{1,2}, Li-li Huang^{1,2}, Bo Yu^{1,2*}

1 Environmental Horticulture Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, Guangdong, China, **2** Guangdong Key Lab of Ornamental Plant Germplasm Innovation and Utilization, Key Laboratory of Urban Agriculture in South China, Ministry of Agriculture, Guangzhou, China

* yubo423@163.com



OPEN ACCESS

Citation: Liu X-f, Sun Y-b, Zhu G-f, Huang L-l, Yu B (2022) Complete chloroplast genomes and comparative analyses of *Hippeastrum* 'milady', *Hippeastrum albertii* and *Hippeastrum reticulatum* (Amaryllidaceae). PLoS ONE 17(8): e0271335. <https://doi.org/10.1371/journal.pone.0271335>

Editor: Shilin Chen, Chinese Academy of Medical Sciences and Peking Union Medical College, CHINA

Received: November 3, 2021

Accepted: June 29, 2022

Published: August 5, 2022

Copyright: © 2022 Liu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The GenBank accession no. of *Hippeastrum* 'Milady' (MT162609), *H. albertii* (MT701522) and *H. reticulatum* (MT701523) were all available from the NCBI database.

Funding: This study is funded by the Key-Area Research and Development Program of Guangdong Province (2020B020220005, 2018B020202002), Guangzhou Science and Technology Project (201807010016,

Abstract

Hippeastrum is a genus of ornamental plants with large, brightly colored flowers. Due to the very high seed-setting rate of the hybridization of *Hippeastrum*, the large population of hybrid progeny and the existence of superparent inheritance, it is difficult to trace the origin of the varieties collected from the market during breeding. In this study, we analyzed the chloroplast genomes of *Hippeastrum* 'Milady', *H. albertii*, and *H. reticulatum* using the Illumina NovaSeq sequencing platform and generated full-length sequences of 158,067, 158,067, and 158,522 bp, respectively. All three genomes had the typical tetrad structure. The large single copy, small single copy, and inverted repeat regions of *H. reticulatum* were observed to be respectively 277, 138, and 20 bp longer than the corresponding regions of *H. 'Milady'* and *H. albertii*. The results of comparative analysis of simple sequence repeats (SSRs), Ka/Ks ratios, codon preferences, and complete sequences of chloroplasts of these three taxa and 14 other plant species were as follows. First, the chloroplast genomes of *H. 'Milady'*, *H. albertii*, and *H. reticulatum* contain 209, 209, and 211 SSR sites, respectively, most of which (123, 123, and 122, respectively) are single nucleotide repeats. Second, leucine, arginine, and serine are the most frequently used amino acids in the three chloroplast genomes. Third, *H. 'Milady'*, *H. albertii*, and *H. reticulatum* are more closely related to *Lycoris* and *Narcissus* than to *Allium* and *Agapanthus*. Our results will provide information on the study of origins or relatedness of native species, and the identification of cultivars.

Introduction

Hippeastrum Herb. (Amaryllidaceae) is a genus of flowering perennials native to South America. The large, brightly colored flowers, which are borne on 50–70-cm long peduncles, can appear at the same time as the leaves. *Hippeastrum* is of high ornamental value, and more than 300 varieties have been developed [1]. In addition, the plants are rich in alkaloids and have thus been investigated for potential anti-anxiety, anti-convulsant, anti-depressive, and anti-

201604020031), Guangdong provincial Science and Technology Department (2018A050506054, 2016LM3169, 2014B070706016). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

toxic properties and for the development of related products [2–4]. Ploidy levels in *Hippeastrum* range from diploid to octoploid [5], and the seed set rate of hybrids is high. Researchers in the Netherlands, the United States, South Africa, and other countries have therefore carried out hybrid breeding work and selected many varieties for use as cut flowers, potted plants, and garden ornamentals [5, 6]. The large number of selected varieties is increasingly hindering the ability to identify and distinguish the original species, hybrid progeny, and cultivars of *Hippeastrum*, which must be performed on the basis of morphology given the lack of genomic data, and therefore it is increasing the difficulty of parental selection in future breeding. Research on the phylogenetic relationships, genetic diversity, and genetic background of *Hippeastrum* and the development of more identification methods and better breeding protocols are thus urgently needed.

Apart from a genetic structural analysis of germplasm resources based on microsatellite markers developed from transcriptome data [5], few studies have been carried out on the genetic structure and phylogeny of *Hippeastrum*. Nuclear, chloroplast, and mitochondrial genomes are widely exploited in phylogenetic analyses. Because of its simple structure, strong conservation, and amenability to sequencing, the chloroplast genome has been used in genetic structural and phylogenetic analyses of a wide range of plant taxa, including *Chlorophytum comosum* [7], Monsteroideae [8], *Hosta* [9], *Aloidendron* [10], *Allium* [11], *Astelia pumila* [12], Agavoideae [13], *Fritillaria ussuriensis* [14], *Amomum kravanh* [15], *Talinum paniculatum* [16], and buckwheat [17]. The chloroplast genomes of various species of Amaryllidaceae have also been published [11, 18–22]. Thus far, however, no research reports have appeared on *Hippeastrum* chloroplast genomes.

In this study, we carried out Illumina sequencing of the chloroplast genomes of *H.* ‘Milady’, *H. alberti*, and *H. reticulatum* and compared their structural features. Using the sequenced genomes, we also analyzed the phylogenetic relationships of these three taxa with Amaryllidaceae and other monocots. The results of our comparative analysis should lay a foundation for future resource evaluation and phylogenetic analysis of the genus.

Materials and methods

Plant materials and extraction of genomic DNA

Genomic DNA was extracted from young leaves (100 mg) of *H.* ‘Milady’, *H. alberti*, and *H. reticulatum* plants using a plant genomic DNA extraction kit (DP305; Tiangen Biotech, Beijing, China). After mechanical (ultrasound) fragmentation, the extracted DNA was subjected to purification, end repair, poly(A) tail addition, and ligation of sequencing adapters and then analyzed by agarose gel electrophoresis for fragment size selection. PCR amplification was performed to generate a sequencing library, and the qualified library was sequenced on the Illumina NovaSeq platform with a sequencing read length of PE150.

Genome sequencing, assembly, and annotation

The paired-end Illumina raw reads were filtered using Trimmomatic [23] and then mapped to the chloroplast genome of the reference species *Xanthorrhoea preissi* (GenBank accession no. NC_035996.1), with Bowtie2 v2.2.4 [24] used to exclude reads of nuclear or mitochondrial origin. SPAdes 3.6.1 [25] and Sequencher 5.3.2 (Gene Codes Inc., Ann Arbor, MI, USA) were used for *de novo* assembly to reconstruct the chloroplast genomes. A “genome walking” technique was adopted to remove gaps [26]. Jellyfish v.2.2.3 [27] was used to correct misassembled contigs. CpGAVAS [28] was used for annotation of the chloroplast genomes, and a circular representation was generated with OGDRAW [29].

Ka/Ks analysis

Base variations resulting in amino acid changes are called non-synonymous mutations, whereas those that do not are termed synonymous mutations. Non-synonymous mutations are generally affected by natural selection. The ratio of the non-synonymous mutation rate (Ka) to the synonymous mutation rate (Ks) indicates the selection effect; a value greater than 1 indicates that a gene is subject to positive selection, while a value less than 1 corresponds to the existence of purifying selection. We aligned gene sequences in mafft v7.310 [30] and then calculated gene Ka/Ks values using KaKs_Calculator v2.0 [31].

Repeat sequence analysis

The Perl script MISA [32] was used to detect microsatellites (mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats) with the following unit size and minimum repeat thresholds: 10 repeat units for mononucleotide SSRs, 5 for dinucleotide SSRs, 4 for trinucleotide SSRs, and 3 each for tetra-, penta-, and hexanucleotide SSRs.

Codon usage and nucleotide diversity

For the identification of codon usage patterns, we used all CDSs present in the three *Hippeastrum* chloroplast genomes to estimate codon usage in CodonW with translational table = 11 [33].

The complete chloroplast genomes of five related species were downloaded from the NCBI database: *Lycoris radiata* (MN158120.1), *Narcissus poeticus* (NC_039825.1), *Agapanthus coddii* (NC_035971.1), *Agave* sp. Pires 2011 (KX931464.1), and *Agave attenuata* (NC_032696.1). The chloroplast genome sequences of these five species and *H. 'Milady'*, *H. alberti*, and *H. reticulatum* were first aligned in MAFFT v7 and then manually adjusted using BioEdit software. A sliding window analysis was then conducted to evaluate the nucleotide variability (π) of the chloroplast genome using DnaSP v5.1 [34]. The step size was set to 200 bp, and the window length was set to 600 bp.

IR contraction and expansion

The chloroplast genome is a circular structure, and IR has four boundaries with LSC and SSC, namely LSC-IRb, IRb-SSC, SSC-IRa and IRa-LSC. During genome evolution, IR boundaries expand and contract, allowing certain genes to enter IR regions or single-copy regions. IRscope (<https://irscope.shinyapps.io/irapp/>) was used for visualizing the genes' differences on the boundaries of the junction sites of the eight chloroplast genomes: *Hippeastrum 'Milady'* (MT162609), *H. alberti* (MT701522), *H. reticulatum* (MT701523), *Lycoris radiata* (MN158120.1), *Narcissus poeticus* (NC_039825.1), *Agapanthus coddii* (NC_035971.1), *Agave* sp. Pires 2011 (KX931464.1) and *Agave attenuata* (NC_032696.1).

Phylogenetic analysis

A phylogenetic tree were constructed by the maximum likelihood method using entire chloroplast genomes. The maximum likelihood analysis was performed using RAXML-HPC BlackBox v.8.1.24 [35] at the CIPRES Science Gateway website based on the best-fit model of evolution (GTR + G) with 1,000 bootstrap replicates. The GenBank accession numbers of the analyzed plant genomes are as follows: *Hippeastrum 'Milady'* (MT162609), *H. alberti* (MT701522), *H. reticulatum* (MT701523), *H. vittatum* (NC_052724.1), *H. rutilum* (MT937175.1), *Lycoris radiata* (MN158120.1), *Narcissus poeticus* (NC_039825.1), *Agapanthus coddii* (NC_035971.1), *Agave* sp. Pires 2011 (KX931464.1), *Agave attenuata* (NC_032696.1),

Allium cepa (KF728080.1), *Asparagus officinalis* (KY364194.1), *Cordyline indivisa* (KX822776.1), *Milla biflora* (KX822778.1), *Aphyllanthes monspeliensis* (KX790360.1), *Aloe vera* (KX377524.1), *Iris sanguinea* (KT626943.1), *Dendrobium bellatulum* (MG595965.1), *Lilium lankongense* (MK757466.1), *Nomocharis pardanthina* (NC_038193.1), *Tulipa altaica* (NC_044780.1), *Arabidopsis thaliana* (NC_000932.1), and *Celosia cristata* (MK470118.1). The latter two species were used as outgroups.

Results

Characteristics of *Hippeastrum* chloroplast sequences

The assembled chloroplast genome sequences of *H. 'Milady'* (GenBank accession no. MT162609) and *H. alberti* (MT701522) were both 158,067 bp, which was 455 bp smaller than that of *H. reticulatum* (MT701523; 158,522 bp). All three sequences had the same GC content (37.93%) and the classical quadripartite structure, namely, a large single copy (LSC) region, a small single copy (SSC) region, and a pair of inverted repeat (IR) regions (Table 1 and Fig 1). The shared sequence structure and GC content of these three taxa suggest that the chloroplast genome is highly conserved in the genus *Hippeastrum*. The three *Hippeastrum* chloroplast genomes were all predicted to encode 133 genes: 86 protein-coding genes (PCGs), 38 transfer RNA (tRNA) genes, 8 ribosomal RNA (rRNA) genes, and 1 pseudogene (Table 1).

Ka/Ks ratios of species pairs

We compared the Ka/Ks values of 86 PCGs in the chloroplast genomes of *H. 'Milady'* or *H. alberti* with those of *H. reticulatum*. The Ka/Ks ratios of *ndhF* and *ndhD* in the first two taxa were greater than 1, namely, 1.78 and 1.87, respectively (S1 Table), which suggests the occurrence of positive selection along these lineages. In contrast, the Ka/Ks ratios of 15 PCGs were less than 1 (S1 Table), indicative of purifying selection. Six of these genes (*rpoC1*, *psbD*, *ycf3*, *rps4*, *petD*, and *rpoA*) had Ka/Ks ratios ranging from 0.1 to 0.3, which implies strong purifying selection (S1 Table).

Analysis of sequence repeats

Numerous microsatellites (simple sequence repeats; SSRs) were detected in the three *Hippeastrum* chloroplast genomes, ranging from 209 in *H. 'Milady'* and *H. alberti* (S2 and S3 Tables) to 211 in *H. reticulatum* (S4 Table). The most abundant SSRs were mono-nucleotide repeats, whose proportions relative to the total number of chloroplast genome SSRs were similar among *H. 'Milady'*, *H. alberti*, and *H. reticulatum*: 58.9% (123), 58.9% (123), and 57.8% (122), respectively (Fig 2a and 2b). Relative to the total number of SSRs, the proportion of mono-nucleotides repeated 11 to 15 times was quite different among taxa, namely, 8.6%, 8.6%, and 4.7% in *H. alberti*, *H. 'Milady'*, and *H. reticulatum*, respectively (Fig 2a and 2b). We also analyzed the distribution of SSRs in different regions of the chloroplast genomes of the three

Table 1. Feature of complete chloroplast genomes of *Hippeastrum* 'Milady', *H. albertii* and *H. reticulatum*.

Taxon	Full		LSC		SSC		IR		Gene number	PCGs	tRNAs	rRNAs	pseudo
	Length (bp)	GC (%)											
<i>H. 'Milady'</i>	158,067	37.93	86,166	35.96	18,271	32.2	26,815	43.05	133	86	38	8	1
<i>H. albertii</i>	158,067	37.93	86,166	35.96	18,271	32.2	26,815	43.05	133	86	38	8	1
<i>H. reticulatum</i>	158,522	37.93	86,443	35.98	18,409	32.16	26,835	43.05	133	86	38	8	1

<https://doi.org/10.1371/journal.pone.0271335.t001>

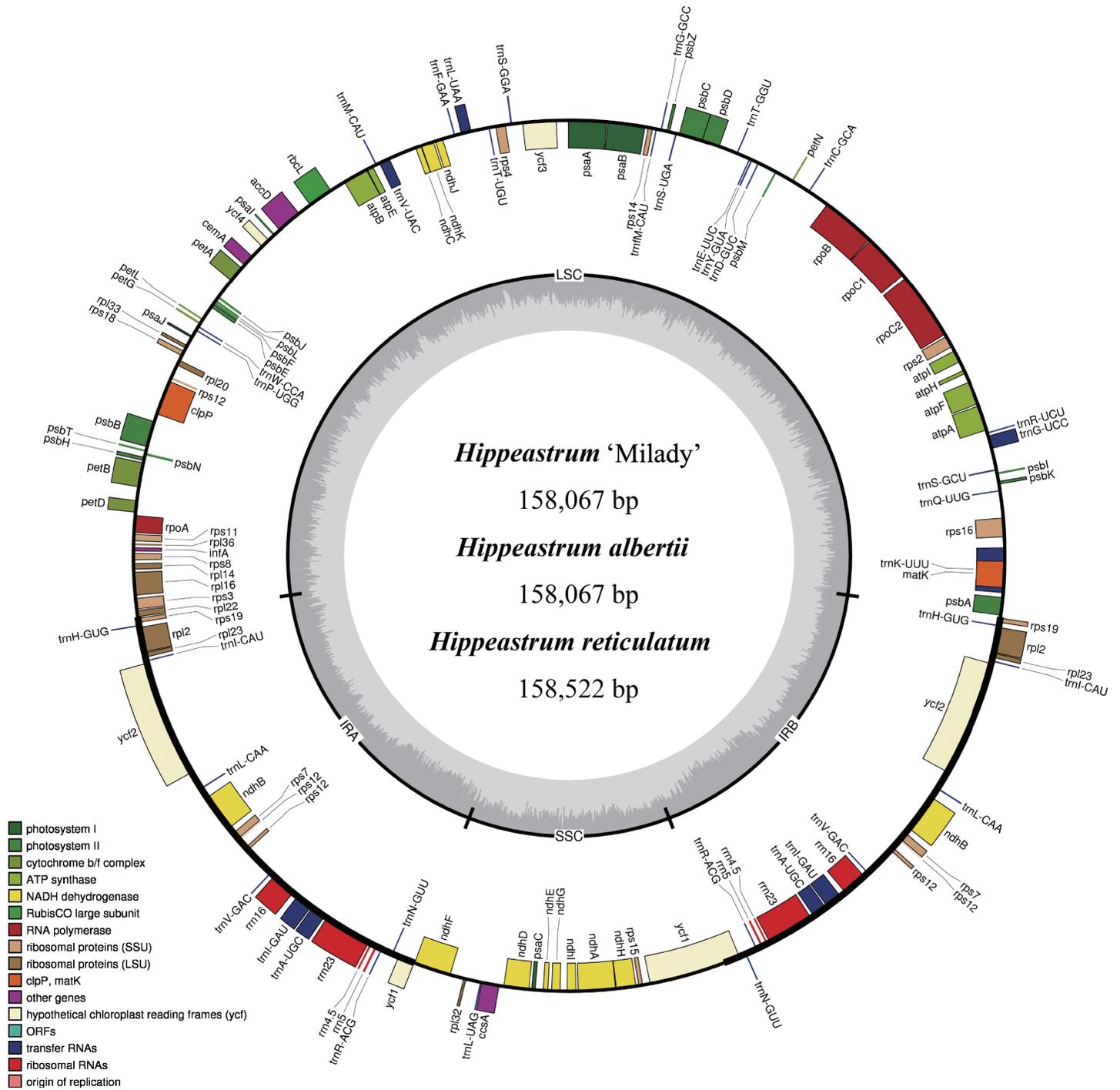


Fig 1. Gene maps of *Hippeastrum* 'Milady', *H. albertii*, and *H. reticulatum*. Genes lying outside the circle are transcribed in a clockwise direction, whereas genes on the inside are transcribed in a counterclockwise direction. Different colors denote known functional groups. The relative GC and AT contents of genomic regions are respectively represented in the inner circle by dark and light gray. LSC, SSC, and IR indicate large single copy, small single copy, and inverted repeat regions, respectively.

<https://doi.org/10.1371/journal.pone.0271335.g001>

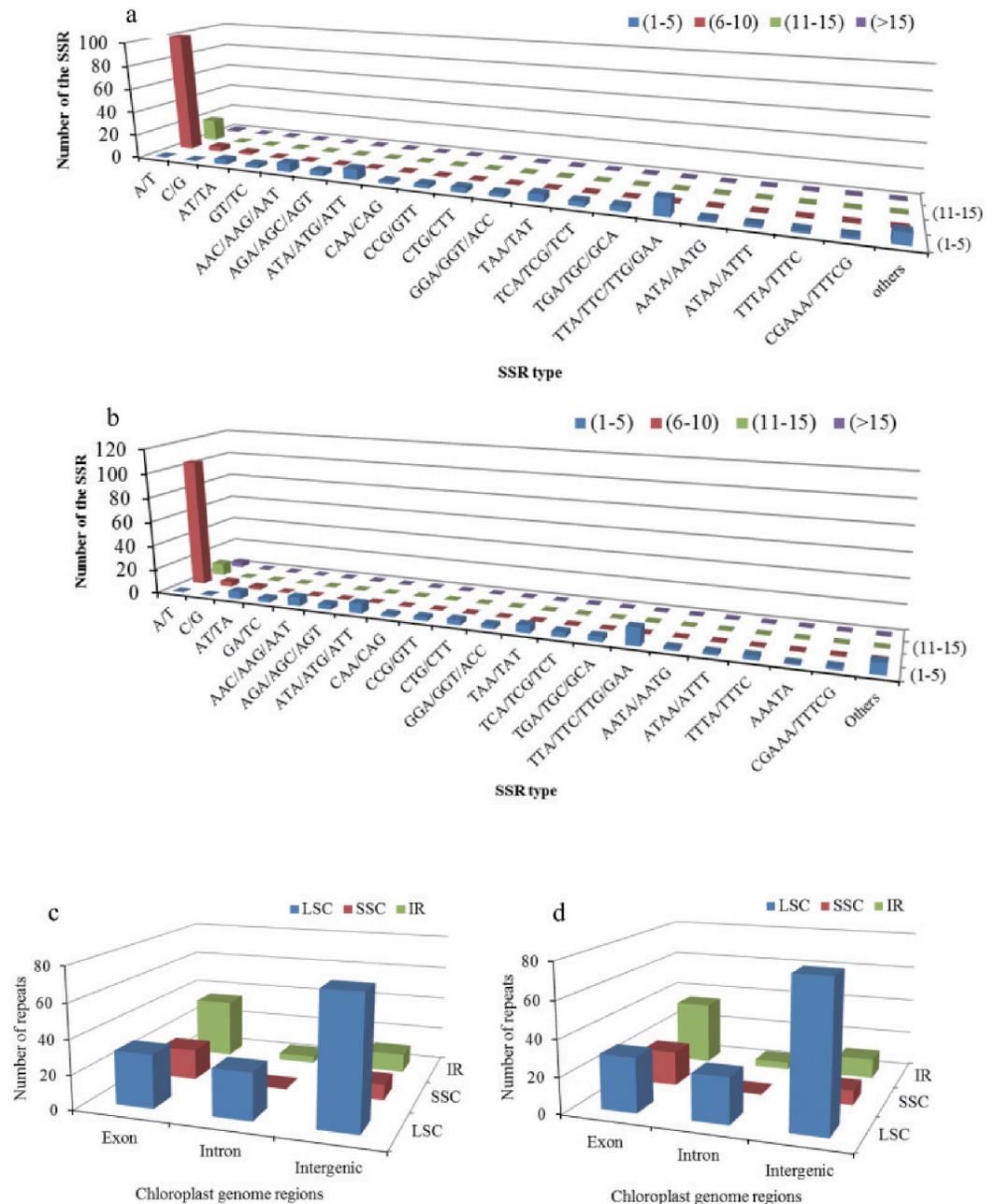


Fig 2. Simple sequence repeats (SSRs) in chloroplast genomes of three *Hippeastrum* taxa. (a–b) SSR types and distributions in *H. alberti* (a), *H. 'Milady'* (a), and *H. reticulatum* (b). (c–d) Distribution of SSRs according to genomic region in *H. alberti* (c), *H. 'Milady'* (c), and *H. reticulatum* (d). LSC, SSC, and IR are large single copy, small single copy, and inverted repeat regions, respectively.

<https://doi.org/10.1371/journal.pone.0271335.g002>

Hippeastrum taxa. We found that the number and overall distribution of SSRs in IR regions were the same in the three taxa. In the LSC and SSC regions, the number of SSRs in the chloroplast genomes of the three *Hippeastrum* taxa was slightly different, but their distribution trends were similar (Fig 2c and 2d).

Codon usage and nucleotide diversity

Codon usage rates vary substantially in the genomes of different species and different organisms. The nonrandom usage of synonymous codons, termed relative synonymous codon usage (RSCU), is considered to be the comprehensive result of natural selection, species mutation, and genetic drift. The three *Hippeastrum* chloroplast genomes all contained 66 codons and a total of 26,390 codons, and their codon usage patterns were also the same. Leucine, arginine, and serine, the most prevalent amino acids, were encoded by the most codons (six), whereas tryptophan was encoded by the fewest. We identified 32 preferred (RSCU > 1.00) and 33 non-preferred (RSCU < 1.00) codons in the three *Hippeastrum* taxa (S5 Table).

Nucleotide diversity values ranged from 0.00000 to 0.04567, with a mean value of 0.01355. Nucleotide diversity ranged from 0.00000 to 0.01837, with an average of 0.0095, in the IR regions, and from 0.0000 to 0.04642, with an average of 0.01372, in the SSC region (S6 Table). Six genes (*ycf1*, *rpl22*, *rps15*, *matK*, *ndhF*, and *ndhD*) had nucleotide diversity values higher than 0.03550, and all except the last two had values above 0.04200. More than 100 mutations were detected in nine genes: *ycf1*, *rpoC2*, *ndhF*, *matK*, *ycf2*, *rpoB*, *ndhD*, *rpoC1*, and *accD* (S6 Table and S1 Fig).

Global genome alignment and analysis

Analysis of SNPs and insertions/deletions (indels) in the chloroplast genome of *H. 'Milady'* or *H. alberti* relative to that of *H. reticulatum* revealed 328 SNPs in *H. 'Milady'* and *H. alberti*, including 131 located in protein-coding regions (S7 Table). We detected 87 indels, consisting of 43 insertions and 41 deletions, in the chloroplast genome of *H. 'Milady'* or *H. alberti* relative to that of *H. reticulatum* (S8 Table). Of these 87 indels, 36 (36.08%) were single-base indels, corresponding to 19 insertions and 17 deletions (S8 Table). We used the SNP sites with Id numbers 14/15, 20, and 21 in the S7 Table to design primers as follows by primer3: No 14 / 15F-TAAGTTCCCATTACGACCC, No 14 / 15R-CCCTACCTTATTGACCGCAA, No 20F-CGACCGAATCGATCAAGAAT, No 20R-TTGGTCTCAACCGTACAGGA, and No 21F-TCCTGTACGGTTGAGACCAA, No 21R-TAGGGCCTTCTGGTTCTTCA. After PCR amplification, Sanger sequencing was performed, and the sequences of the three species were compared. The results found that the sequences of *H. 'Milady'* and *H. alberti* are identical, and both of them are different from *H. reticulatum* at the SNP site (S2-S4 Figs.).

IR contraction and expansion

Detailed comparisons of the four junctions LSC-IRb (JLB), IRb-SSC (JSB), SSC-Ira (JSA), IRa-LSC (JLA), and among six Amaryllidaceae chloroplast genomes (*Hippeastrum 'Milady'*, *H. alberti*, *H. reticulatum*, *Lycoris radiata*, *Narcissus poeticus*, and *Agapanthus coddii*) and two Asparagaceae chloroplast genomes (*Agave* sp. Pires 2011 and *Agave attenuata*) are presented in Fig 3. IR region of three *Hippeastrum* species chloroplast genomes was highly conserved, and slightly structure variation were found in the JLB and JSB regions (Fig 3). The *rpl22-rps19* gene were located at the junctions of the JLB regions in *H. 'Milady'*, *H. alberti*, *H. reticulatum*, *L. radiata*, *N. poeticus*, *A. coddii*, and *A. sp. Pires 2011*, and *A. attenuata*, and the *rps19* gene is located in the IRb region 52, 49, and 52 bp away from the JLB border in three *Hippeastrum* chloroplast genomes, respectively (Fig 3). The *trnN-ndhF* genes were located in the junctions of the JSB regions. The *trnN* gene is located in the IRb region ~1300 bp away from the JSB border in eight species except *N. poeticus* (3086 bp) (Fig 3). The *ycf1* gene was located in the junctions of the JSA regions in the eight species (Fig 3). However, the distance (2609 bp) from the JSA boundary in the part of the *ycf1* gene located in the SSC region in the chloroplast genome of *N. poeticus* as significantly different (about 4400 bp) from the other seven species (Fig 3).

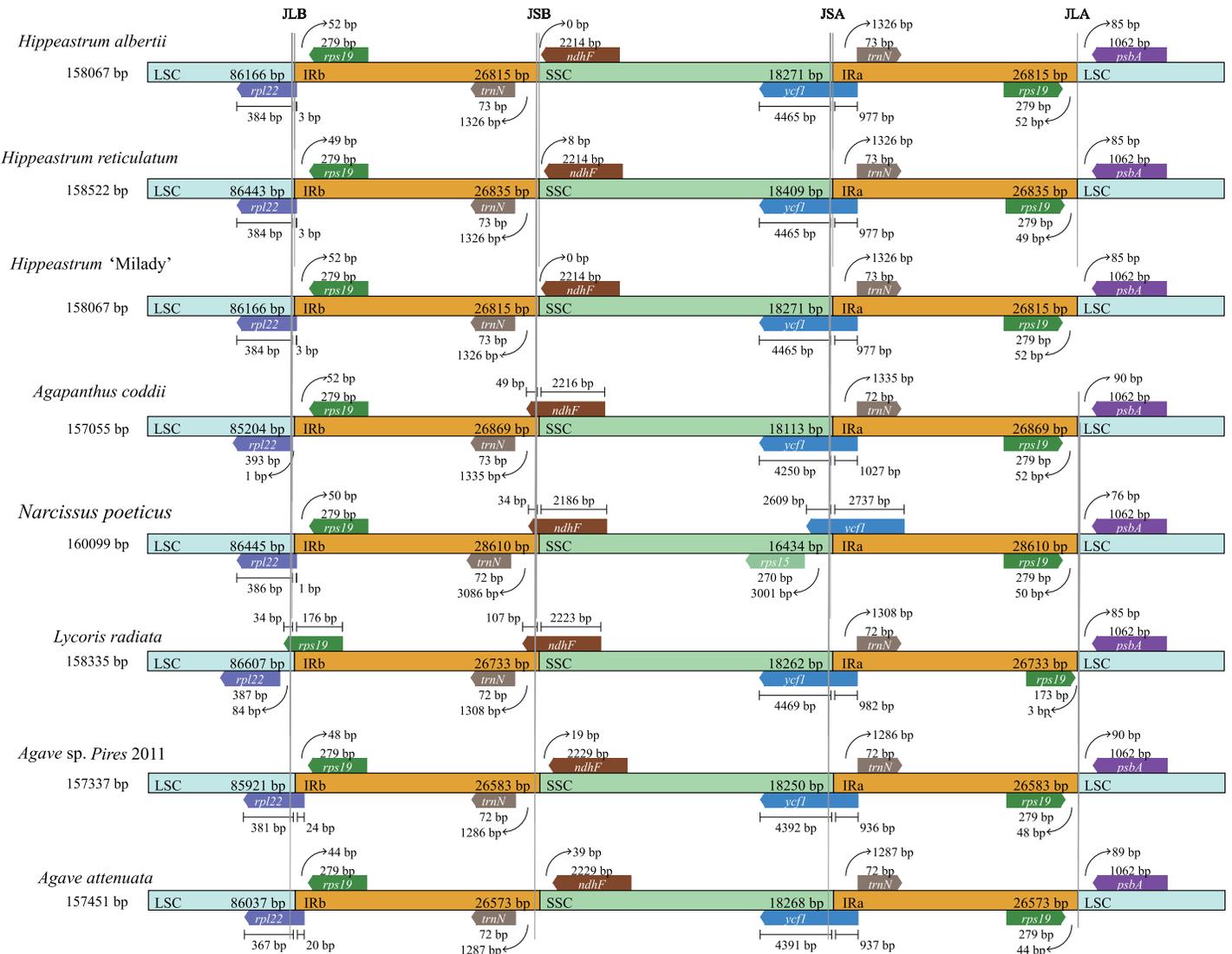


Fig 3. Comparison of the LSC, SSC, and IR regions among eight chloroplast genomes. Boxes above the main line indicate the adjacent border genes. The figure is not to scale with respect to sequence length, and shows only relative changes at or near the IR/SSC borders.

<https://doi.org/10.1371/journal.pone.0271335.g003>

The *rps19-psbA* genes were located in the junctions of JLA regions, and the *rps19* gene was located in the IRa region 3 bp away from the JLB border in *L. radiata* different from the other 7 species (Fig 3).

Phylogenetic analysis

To analyze phylogenetic relationships of *Hippeastrum*, we generated a multiple alignment of full-length sequences of chloroplast genomes of *H. 'Milady'*, *H. albertii*, *H. reticulatum*, and 18 other monocots plus two outgroups (*Arabidopsis thaliana* and *Celosia cristata*) and carried out a maximum likelihood analysis. In the resulting tree shown in Fig 4, *H. albertii* and *H. 'Milady'*, which have identical chloroplast genome sequences, are clustered together, and the five genomes of *Hippeastrum* form a distinct clade (Fig 4). All nine analyzed species in Amaryllidaceae group together, and *H. 'Milady'*, *H. albertii*, and *H. reticulatum* are more closely related to

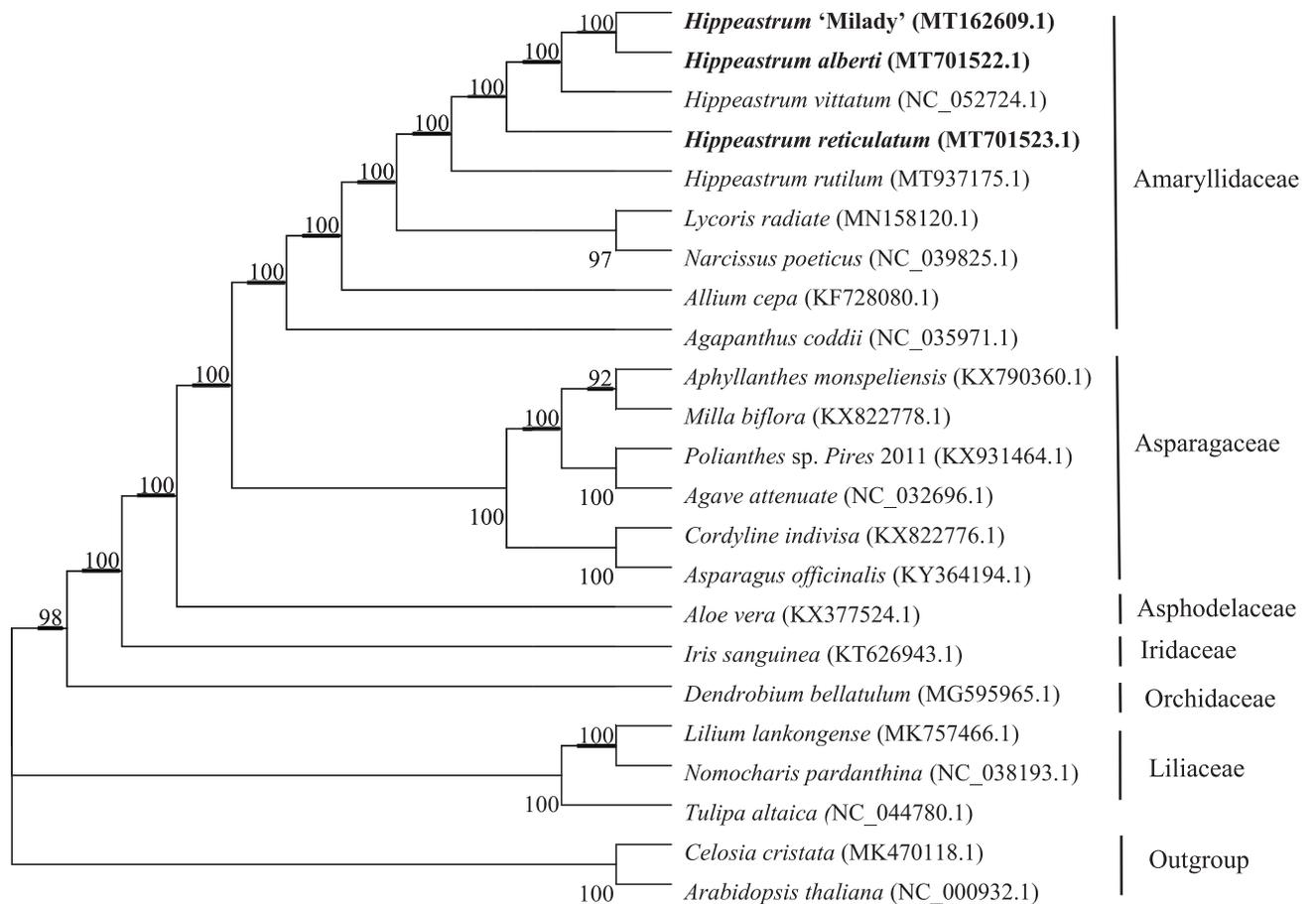


Fig 4. Phylogenetic tree based on maximum likelihood analysis of sequences from whole chloroplast genomes of 18 plant species. *Arabidopsis thaliana* and *Celosia cristata* were used as outgroups. Numbers at nodes are bootstrap support percentages.

<https://doi.org/10.1371/journal.pone.0271335.g004>

H. vittatum than to *H. rutilum*. According to the tree, *H. 'Milady'*, *H. alberti*, and *H. reticulatum* are more closely related to *Lycoris radiata* and *Narcissus poeticus* than to *Allium cepa* and *Agapanthus coddii* (Fig 4). The phylogenetic relationships of Amaryllidaceae to Asparagaceae, Asphodelaceae, Iridaceae, and Orchidaceae revealed in the tree should provide a theoretical basis for future evolutionary analyses of Amaryllidaceae species.

Discussion

In this study, the complete chloroplast genomes of *H. 'Milady'*, *H. alberti*, and *H. reticulatum* obtained by Illumina sequencing ranged from 158.1 to 158.5 kb in length. Comparative analysis revealed that the number of PCGs and tRNAs in the three sequenced genomes is similar to that of other species in Amaryllidaceae, including *Lycoris* (87 PCGs) [22] and *Allium* (81–87 PCGs) [36], but exceeds that of *Celosia cristata* (73 PCGs) [37]. The number of rRNAs in the chloroplast genomes, eight, is conserved in Amaryllidaceae.

Our analysis of Ka/Ks ratios uncovered two genes undergoing positive selection in *H. 'Milady'*/*H. alberti* relative to *H. reticulatum*: *ndhD* and *ndhF*, both of which are single-copy genes located in the SSC region (Fig 1). We also selected *Lycoris radiata*, *Narcissus poeticus*, and *Agapanthus coddii* from Amaryllidaceae as well as *Agave* sp. Pires 2011 and *Agave attenuata* from Asparagaceae for a Ka/Ks ratio pairwise analysis. The results of this analysis suggest

that the number of positively selected genes between species in different families was greater than that between species in the same family (S9–S18 Tables). These findings are in agreement with the results of previous research on *Chrysosplenium* showing lower Ka/Ks ratios at the chloroplast genome level within *Chrysosplenium* compared with non-*Chrysosplenium* species [38], but they differ from the outcome of a study in *Allium* in which higher pairwise Ka/Ks ratios were observed in *Allium* (Allioideae) species pairs than in non-Allioideae species pairs [11]. Purifying selection constantly sweeps away deleterious mutations in a population. The lower number of positively selected chloroplast genes in Amaryllidaceae compared with Asparagaceae may be the evolutionary result of the preservation of the adaptive characteristics of Amaryllidaceae species.

The GC content of plant species ranged from 19.5% to 42.1% [39]. In this study, we observed a lower GC content in the chloroplast genomes of *H. ‘Milady’*, *H. albertii*, and *H. reticulatum*, which is similar to *Lycoris* [22] and *Allium* [36]. AT-mutation pressure or AT-biased gene conversion translational efficiency may lead to the paucity of G and C nucleotides observed in plastid genomes [40, 41]. Because they have three hydrogen bonds, GC pairs are more stable than AT pairs, which have only two hydrogen bonds [42].

Expansion and contraction at the junction regions of IRs are the main explanations for the size variation and are common evolutionary events among chloroplast genomes in different species [43–45]. In this study, we compared three *Hippeastrum* species with two species of other genus of Amaryllidaceae and two species of Asparagaceae using chloroplast genomes. It was found that the expansion of IR regions resulted in the IR regions of *Narcissus poeticus* being more than a kilo base longer than the other seven species sequences selected. At the same time, we also found that the distance between *rps19* gene and JLA was shortened in JLA regions. These various regions in the chloroplast genome may provide us with more information in genetic structural analysis. Our phylogenetic analysis clearly revealed that 6 species of Amaryllidaceae clustered well into one clade and 5 *Hippeastrum* species clustered well into one clade.

Supporting information

S1 Fig. The nucleotide diversity value of all genes in *Hippeastrum ‘Milady’*, *H. albertii*, *H. reticulatum*, *Lycoris radiata*, *Narcissus poeticus*, *Agapanthus coddii*, *Agave* sp. *Pires* 2011 and *Agave attenuate*.

(EPS)

S2 Fig. Sequence alignment results after PCR amplification and sequencing of primer No 14/15.

(TIF)

S3 Fig. Sequence alignment results after PCR amplification and sequencing of primer No 20.

(TIF)

S4 Fig. Sequence alignment results after PCR amplification and sequencing of primer No 21.

(TIF)

S1 Table. Ka/Ks ratios of the chloroplast genomes between *H. ‘Milady’/H. albertii* and *H. reticulatum*.

(XLSX)

- S2 Table. Information about SSRs in the chloroplast genome of *Hippeastrum* ‘Milady’.**
(XLSX)
- S3 Table. Information about SSRs in the chloroplast genome of *Hippeastrum albertii*.**
(XLSX)
- S4 Table. Information about SSRs in the chloroplast genome of *Hippeastrum reticulatum*.**
(XLSX)
- S5 Table. Statistical results of relative synonymous codon usage in the chloroplast genome of *H.* ‘Milady’, *H. albertii* and *H. reticulatum*.**
(XLT)
- S6 Table. Statistical results of the nucleotide diversity value.**
(XLSX)
- S7 Table. SNPs in the chloroplast genome of *H.* ‘Milady’ or *H. albertii* relative to that of *H. reticulatum*.**
(XLSX)
- S8 Table. Indels in the chloroplast genome of *H.* ‘Milady’ or *H. albertii* relative to that of *H. reticulatum*.**
(XLSX)
- S9 Table. Ka/Ks ratios of the chloroplast genomes between *H.* ‘Milady’/*H. albertii* and *Agave* sp. *Pires* 2011.**
(XLSX)
- S10 Table. Ka/Ks ratios of the chloroplast genomes between *H.* ‘Milady’/*H. albertii* and *Lycoris radiate*.**
(XLSX)
- S11 Table. Ka/Ks ratios of the chloroplast genomes between *H.* ‘Milady’/*H. albertii* and *Agave attenuate*.**
(XLSX)
- S12 Table. Ka/Ks ratios of the chloroplast genomes between *H.* ‘Milady’/*H. albertii* and *Agapanthus coddii*.**
(XLSX)
- S13 Table. Ka/Ks ratios of the chloroplast genomes between *H.* ‘Milady’/*H. albertii* and *Narcissus poeticus*.**
(XLSX)
- S14 Table. Ka/Ks ratios of the chloroplast genomes between *H. reticulatum* and *Agave* sp. *Pires* 2011.**
(XLSX)
- S15 Table. Ka/Ks ratios of the chloroplast genomes between *H. reticulatum* and *Lycoris radiate*.**
(XLSX)
- S16 Table. Ka/Ks ratios of the chloroplast genomes between *H. reticulatum* and *Agave attenuate*.**
(XLSX)

S17 Table. Ka/Ks ratios of the chloroplast genomes between *H. reticulatum* and *Agapanthus coddii*.

(XLSX)

S18 Table. Ka/Ks ratios of the chloroplast genomes between *H. reticulatum* and *Narcissus poeticus*.

(XLSX)

Acknowledgments

We thank Liwen Bianji (Edanz) (www.liwenbianji.cn/ac) for editing the English text of a draft of this manuscript.

Author Contributions

Conceptualization: Xiao-fei Liu.

Data curation: Xiao-fei Liu, Li-li Huang.

Investigation: Ying-bo Sun.

Methodology: Xiao-fei Liu, Gen-fa Zhu.

Writing – original draft: Xiao-fei Liu.

Writing – review & editing: Bo Yu.

References

1. Jamil MK, Rahman MM, Hossain MM, Hossain MT, Karim AS. Influence of sucrose and aluminium sulphate vase life of cut *Hippeastrum* flower (*Hippeastrum hybridum* Hort.) as influenced. Bangladesh J Agri Res. 2016, 41: 221–234.
2. De Andrade JP, Pigni NB, Torras-Claveria L, Guo Y, Berkov S, Reyes-Chilpa R, et al. Alkaloids from the *Hippeastrum* genus: chemistry and biological activity. Revista Latinoamericana de Química. 2012, 40: 83–98.
3. Nair JJ, Bastida J, Codina C, Viladomat F, van Staden J. Alkaloids of the South African Amaryllidaceae: a review. Nat. Prod. Commun. 2013, 8: 1335–1350. PMID: [24273880](https://pubmed.ncbi.nlm.nih.gov/24273880/)
4. Nair JJ, van Staden J. Antiplasmodial constituents in the minor alkaloid groups of the Amaryllidaceae. S Afr J Bot. 2019, 126: 362–370.
5. Wang Y, Chen D, He X, Shen J, Xiong M, Wang X, et al. Revealing the complex genetic structure of cultivated amaryllis (*Hippeastrum hybridum*) using transcriptome-derived microsatellite markers. Sci Rep. 2018, 8: 10645. <https://doi.org/10.1038/s41598-018-28809-9> PMID: [30006536](https://pubmed.ncbi.nlm.nih.gov/30006536/)
6. Vazquez C, Reed ST, Dunn C. Nitrogen Fertilization as Ammonium or Nitrate-N on *Hippeastrum hybridum* BulbGrowth. Agr Sci. 2015, 6: 1547–1554.
7. Munyao JN, Dong X, Yang JX, Mbandi EM, Wanga VO, Oulo MA, et al. Complete Chloroplast Genomes of *Chlorophytum comosum* and *Chlorophytum gallabatense*: Genome Structures, Comparative and Phylogenetic Analysis. Plants (Basel). 2020, 9: 296. <https://doi.org/10.3390/plants9030296> PMID: [32121524](https://pubmed.ncbi.nlm.nih.gov/32121524/)
8. Henriquez CL, Abdullah AI, Carlsen MM, Zuluaga A, Croat TB, McKain MR, et al. Molecular evolution of chloroplast genomes in Monsteroideae (Araceae). Planta. 2020, 251: 72. <https://doi.org/10.1007/s00425-020-03365-7> PMID: [32112137](https://pubmed.ncbi.nlm.nih.gov/32112137/)
9. Lee SR, Kim K, Lee BY, Lim CE. Complete chloroplast genomes of all six Hosta species occurring in Korea: molecular structures, comparative, and phylogenetic analyses. BMC Genomics. 2019, 20: 833. <https://doi.org/10.1186/s12864-019-6215-y> PMID: [31706273](https://pubmed.ncbi.nlm.nih.gov/31706273/)
10. Malakasi P, Bellot S, Dee R, Grace OM. Museomics Clarifies the Classification of Aloidendron (Asphodelaceae), the Iconic African Tree Aloes. Front Plant Sci. 2019, 10: 1227. <https://doi.org/10.3389/fpls.2019.01227> PMID: [31681358](https://pubmed.ncbi.nlm.nih.gov/31681358/)
11. Xie DF, Yu HX, Price M, Xie C, Deng YQ, Chen JP, et al. Phylogeny of Chinese *Allium* Species in Section Daghestanica and Adaptive Evolution of *Allium* (Amaryllidaceae, Alliioideae) Species Revealed by

- the Chloroplast Complete Genome. *Front Plant Sci.* 2019, 10: 460. <https://doi.org/10.3389/fpls.2019.00460> PMID: 31114591
12. Pfanzelt S, Albach DC, von Hagen KB. Extremely low levels of chloroplast genome sequence variability in *Astelia pumila* (Asteliaceae, Asparagales). *Peer J.* 2019, 7: e6244.
 13. McKain MR, McNeal JR, Kellar PR, Eguiarte LE, Pires JC, Leebens-Mack J. Timing of rapid diversification and convergent origins of active pollination within Agavoideae (Asparagaceae). *Am J Bot.* 2016, 103: 1717–1729. <https://doi.org/10.3732/ajb.1600198> PMID: 27793858
 14. Park I, Kim WJ, Yeo SM, Choi G, Kang YM, Piao R, et al. The complete chloroplast genome sequences of *Fritillaria ussuriensis* Maxim. and *Fritillaria cirrhosa* D. Don and comparative analysis with other *Fritillaria* species. *Molecules* 2017, 22: 982. <https://doi.org/10.3390/molecules22060982> PMID: 28608849
 15. Wu M, Li Q, Hu Z, Li X, Chen S. The complete *Amomum kravanh* chloroplast genome sequence and phylogenetic analysis of the commelinids. *Molecules* 2017, 22: 1875. <https://doi.org/10.3390/molecules22111875> PMID: 29104233
 16. Liu X, Li Y, Yang H, Zhou B. Chloroplast genome of the folk medicine and vegetable plant *Talinum paniculatum* (Jacq.) Gaertn.: Gene organization, comparative and phylogenetic analysis. *Molecules* 2018, 23: 857. <https://doi.org/10.3390/molecules23040857> PMID: 29642545
 17. Wang CL, Ding MQ, Zou CY, Zhu XM, Tang Y, Zhou ML, et al. Comparative analysis of four buckwheat species based on morphology and complete chloroplast genome sequences. *Sci. Rep.* 2017, 7: 6514. <https://doi.org/10.1038/s41598-017-06638-6> PMID: 28747666
 18. Jin SW, Park JY, Kang SJ, Park HS, Shim H, Lee TJ, et al. The complete chloroplast genome sequence of Magic Lily (*Lycoris squamigera*). *Mitochondrial DNA B.* 2018, 3: 1210–1211.
 19. Zhang F, Shu X, Wang T, Zhuang W, Wang ZL. The complete chloroplast genome sequence of *Lycoris radiata*. *Mitochondrial DNA B.* 2019, 4: 2886–2887. <https://doi.org/10.1080/23802359.2019.1660265> PMID: 33365775
 20. Zhang F, Tong H, Yang H, Wang T, Zhuang W, Shu X, et al. Characterisation of the complete chloroplast genome of *Lycoris longituba* (Amaryllidaceae). *Mitochondrial DNA B.* 2019, 4: 3782–3783.
 21. Zhang F, Zhuang W, Shu X, Wang T, Wang ZL. Complete chloroplast genome of *Lycoris sprengeri* (Amaryllidaceae) and genetic comparison. *Mitochondrial DNA B.* 2019, 4: 3577–3578.
 22. Zhang F, Wang T, Shu X, Wang N, Zhuang W, Wang Z. Complete Chloroplast Genomes and Comparative Analyses of *L. chinensis*, *L. anhuiensis*, and *L. aurea* (Amaryllidaceae). *Int J Mol Sci.* 2020, 21: 5729.
 23. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* 2014, 30: 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
 24. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods.* 2012, 9: 357–359. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
 25. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol.* 2012, 19: 455–77. <https://doi.org/10.1089/cmb.2012.0021> PMID: 22506599
 26. Souza UJBd, Nunes R, Targueta CP, Diniz-Filho JAF, Telles MPC. The complete chloroplast genome of *Stryphnodendron adstringens* (Leguminosae—Caesalpinioideae): comparative analysis with related Mimosoid species. *Sci Rep.* 2019, 9: 14206. <https://doi.org/10.1038/s41598-019-50620-3> PMID: 31578450
 27. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics.* 2011, 27: 764–770. <https://doi.org/10.1093/bioinformatics/btr011> PMID: 21217122
 28. Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and Genbank submission of completely sequenced chloroplast genome sequences. *BMC Genomics.* 2012, 13: 715. <https://doi.org/10.1186/1471-2164-13-715> PMID: 23256920
 29. Lohse M, Drechsel O, Bock R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 2007, 52: 267–274. <https://doi.org/10.1007/s00294-007-0161-y> PMID: 17957369
 30. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 2013, 30: 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
 31. Wang D, Zhang Y, Zhang Z, Jiang Z, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom Proteom Bioinf.* 2010, 8: 77–80. [https://doi.org/10.1016/S1672-0229\(10\)60008-3](https://doi.org/10.1016/S1672-0229(10)60008-3) PMID: 20451164

32. MISA-Microsatellite Identification Tool. <http://pgrc.ipk-gatersleben.de/misa/> (accessed on 20 September 2017).
33. Peden JF. Analysis of codon usage. *Univ. Nottm.* 2000, 90: 73–74.
34. Librado P, Rozas J. DnaSPv5: software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009, 25: 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187> PMID: 19346325
35. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006, 22: 2688–2690. <https://doi.org/10.1093/bioinformatics/btl446> PMID: 16928733
36. Xie DF, Tan JB, Yu Y, Gui LJ, Su DM, Zhou SD, et al. Insights into phylogeny, age and evolution of *Allium* (Amaryllidaceae) based on the whole plastome sequences. *Ann. Bot. Lond.* 2020, 125: 1039–1055. <https://doi.org/10.1093/aob/mcaa024> PMID: 32239179
37. Liu X, Ye Y, Liu J, Yu B, Xu Y. Complete chloroplast genome sequence and phylogenetic analysis of *Celosia cristata* 'Xiaguang', Mitochondrial DNA Part B. 2020, 5: 1338–1339.
38. Wu Z, Liao R, Yang T, Dong X, Lan D, Qin R, et al. Analysis of six chloroplast genomes provides insight into the evolution of *Chrysosplenium* (Saxifragaceae). *BMC Genomics.* 2020, 21: 621. <https://doi.org/10.1186/s12864-020-07045-4> PMID: 32912155
39. Smith DR. Unparalleled GC content in the plastid DNA of Selaginella. *Plant Mol. Biol.* 2009, 71: 627–639. <https://doi.org/10.1007/s11103-009-9545-3> PMID: 19774466
40. Khakhlova O, Bock R. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* 2010, 46: 85–94.
41. And KD, Voelker LRL. Molecular biology of mycoplasmas. *Annu. Rev. Microbiol.* 2003, 50: 25–57.
42. Jia Q, Wu H, Zhou X, Gao J, Zhao W, Aziz J. et al. A "GC-rich" method for mammalian gene expression: a dominant role of non-coding DNA GC content in regulation of mammalian gene expression. *Sci China Life Sci.* 2010, 53: 94–100. <https://doi.org/10.1007/s11427-010-0003-x> PMID: 20596960
43. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boorem JL, et al. Comparative chloroplast genomics: Analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genom.* 2007, 8: 174–201. <https://doi.org/10.1186/1471-2164-8-174> PMID: 17573971
44. Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol. Biol.* 2008, 8: 36–50. <https://doi.org/10.1186/1471-2148-8-36> PMID: 18237435
45. Liu XF, Zhu GF, Li DM, Wang XJ. Complete chloroplast genome sequence and phylogenetic analysis of *Spathiphyllum* 'Parrish'. *PLoS ONE* 2019, 14(10): e0224038. <https://doi.org/10.1371/journal.pone.0224038> PMID: 31644545