



OPEN

Complete chloroplast genome of *Lilium ledebourii* (Baker) Boiss and its comparative analysis: lights into selective pressure and adaptive evolution

Morteza Sheikh-Assadi¹✉, Roohangiz Naderi¹✉, Mohsen Kafi¹, Reza Fatahi¹, Seyed Alireza Salami¹ & Vahid Shariati²

Lilium ledebourii (Baker) Boiss is a rare species, which exhibits valuable traits. However, before its genetic diversity and evolutionary were uncovered, its wild resources were jeopardized. Moreover, some ambiguities in phylogenetic relationships of this genus remain unresolved. Therefore, obtaining the whole chloroplast sequences of *L. ledebourii* and its comparative analysis along with other *Lilium* species is crucial and pivotal to understanding the evolution of this genus as well as the genetic populations. A multi-scale genome-level analysis, especially selection pressure, was conducted. Detailed third-generation sequencing and analysis revealed a whole chloroplast genome of 151,884 bp, with an ordinary quadripartite and protected structure comprising 37.0% GC. Overall, 113 different genes were recognized in the chloroplast genome, consisting of 30 distinct tRNA genes, four distinct ribosomal RNAs genes, and 79 unique protein-encoding genes. Here, 3234 SSRs and 2053 complex repeats were identified, and a comprehensive analysis was performed for IR expansion and contraction, and codon usage bias. Moreover, genome-wide sliding window analysis revealed the variability of *rpl32-trnL-ccsA*, *petD-rpoA*, *ycf1*, *psbI-trnS-trnG*, *rps15-ycf1*, *trnR*, *trnT-trnL*, and *trnP-psaJ-rpl33* were higher among the 48 *Lilium* cp genomes, displaying higher variability of nucleotide in SC regions. Following 1128 pairwise comparisons, *ndhB*, *psbJ*, *psbZ*, and *ycf2* exhibit zero synonymous substitution, revealing divergence or genetic restriction. Furthermore, out of 78 protein-coding genes, we found that *accD* and *rpl36* under positive selection: however, at the entire-chloroplast protein scale, the *Lilium* species have gone through a purifying selection. Also, a new phylogenetic tree for *Lilium* was rebuilt, and we believe that the *Lilium* classification is clearer than before. The genetic resources provided here will aid future studies in species identification, population genetics, and *Lilium* conservation.

In keeping with APG IV, Liliales are made up of about 1500 species divided into 10 families that have been categorized by various authors¹. The Liliales order is constantly being studied botanically, as variously amended family circumscriptions²⁻⁴. Some of the changes in Liliales, for example, include⁴⁻⁶: (1) Petermanniaceae has been recognized as a family, while it was within Colchicaceae; (2) Luzuriagaceae has been placed in Alstroemeriaceae; and (3) Corsiaceae has recently been placed in Liliales. Liliaceae contains 15 genera and around 900 species⁷. The family's classification has shifted significantly as a result of modern molecular phylogenetic analysis^{1,8,9}. *Lilium* is a genus in Liliaceae, which contains circa 100 species¹⁰. This genus classification has historically been obscurant. Based on morphological characteristics, several classifications of *Lilium* have been suggested. This genus was classified into seven sections relying on 13 morphological characteristics¹¹, which have primarily been applied to distinguish *Lilium* species and *Lilium* phylogeny exploration¹². Nevertheless, this type of species classification based on morphological traits is oftentimes dynamic and untrustworthy and is frequently influenced by environmental circumstances; that is why there is some disagreement about Comber's classification. Therefore,

¹Department of Horticultural Science, Faculty of Agricultural Science and Engineering, University of Tehran, Karaj, Iran. ²NIGEB Genome Center, National Institute of Genetic Engineering and Biotechnology, Tehran, Iran. ✉email: m.asadi66@ut.ac.ir; maderi@ut.ac.ir

the Comber's classification verification has been checked by molecular phylogenetic investigations^{13–16}. Additionally, Kim et al.¹⁷, Du et al.¹⁸, and Kim et al.¹² evaluated *Lilium* phylogeny using 9, 16, and 28 species, respectively, which somewhat resolved the phylogenetic relationships, but some ambiguities still remain. For example, due to sampling restrictions for *Lophophorum* and *Nomocharis*, the position of *L. distichum* has not been clarified enough. Kim et al.¹², according to their results, pointed out that the position of *L. candidum* is uncertain, and more sampling is required to resolve it.

Furthermore, due to the sampling and hybridization of these species to produce today's lilies, understanding wild lilies is essential for achieving breeding program goals¹⁹. *L. ledebourii*, called Susan-e Chelcheragh (SCH) in Persian, is a rare species in the genus *Lilium*. It has only been seen in Iran and Azerbaijan. Due to uncontrolled grazing and poaching, it is extremely endangered, now being protected only in a small area of Damash village^{20,21}. *L. ledebourii* exhibits valuable traits, including attractive white flower²⁰, a high number of flowers, usually 2–15²² and even up to 24 (personal observations of the author, 2016), the sweet fragrance²³ quoted by²⁴, an excellent vase-life, a vigorous growth, a good tolerance to low light density and low temperatures²⁵. These make it no less beautiful than the commercial species. However, despite being highly valuable and, more importantly, endangered, has not only little effort has been made to use the species in population genetic studies, but also its rightful position among lilies is unclear.

DNA barcoding is one of the most efficient methods for characterizing and classifying various organisms at the species and genus levels²⁶. One of the research hotspots for DNA barcode screening is chloroplast (cp), which can be employed as a super-barcode to solve the classification problem in phylogenetic studies and species identification^{27,28}. Chloroplasts, the energy generators of plant cells, ensure life on the earth^{29,30}. This vital organ serves as a signaling hub in the cell, releasing a diverse range of signals that adjust a well-regulated and proper reaction to any condition³¹. The contents, structure, and gene organization of a chloroplast genome are more strictly conserved than a nucleic genome³². It has a much lower substitution rate than a nucleic genome, and the substitution rate is even lower in two inverted repeat regions³³. Noteworthy, information included in chloroplast genomes, as well as their almost nonrecombinant traits³⁴, maternal transmission³⁵, have made the chloroplast genome a good source for searching for clues about the origins of populations as well as for phylogenetic reconstructions, thereby clearing the ambiguities present in the evolutionary relationships³⁶. Hence, today's strategies for discovering plant molecular phylogeny rely profoundly on cp genome sequences. Additionally, because of advances in next-generation sequencing technology, chiefly third-generation sequencing such as PacBio, leaning on single-molecule real-time (SMRT), which produces reads > 10 kb, the decoding of chloroplast genomes has been accelerated^{37,38}.

Other advantages of chloroplasts include using mutation hotspot sites, and the single sequence repeats to aid population genetics and species identification³⁹. The selection pressure that a species face during evolution is another fascinating aspect of chloroplast genome analysis, which divulged the impact of various environmental pressures on cp genomes when it comes to long-term evolution⁴⁰. Recent studies have discovered a slew of positive selection genes e.g., the *accD* in *Ipomoea*⁴¹, and *petG*, *rpl36*, and *atpB* in *Aquilegia*⁴², as well as purifying selection at cp genome-scale in *Stauntonia*⁴³, but we know very little about it in *Lilium* at the gene-level and nothing at the cp genome-level.

This study, employing PacBio platform, reports the whole chloroplast genome of *L. ledebourii*, a precious endangered species. In addition, using the genome data, we conducted a multi-scale genome-level analysis among this species, newly unemployed (at the genome-scale) species, and other *Lilium* species. In particular, for the first time, we presented a comprehensive analysis of the selection pressure between *Lilium* species at both the gene-level and genome-level. This study covered overlooked topics in previous studies as much as possible. Lastly, employing the richer taxon sampling, we rebuilt a new phylogenetic tree for *Lilium* based on the whole cp genome, and we believe that we have given more resolution to the *Lilium* classification than in earlier studies. This article may be the primary cornerstone for future molecular studies and genetic improvement of *L. ledebourii*.

Results

The chloroplast genome features of *L. ledebourii*. The *L. ledebourii* cp genome, as typical, showed the ordinary quadripartite and protected structure with 151,884 bp in length. The two regions of LSC (81,412 bp) and SSC (17,620 bp) were present in the *L. ledebourii* genome, separated by two inverted repeats (26426 bp), IRA and IRb (Fig. 1). *L. ledebourii* cp chloroplast has 37.0% GC content, with IRs having the highest (42.5%) and SSC having the lowest (37.5%). Overall, 113 different genes were recognized in the SCH chloroplast genome, consisting of 30 distinct tRNA genes, four distinct ribosomal RNAs (4.5S, 5S, 16S, and 23S) genes, and 79 unique protein-encoding genes (Fig. 1, Table 1). Primarily based on their features, all genes are classified into five main categories (Table 1). Among 113 genes, most of the genes happen without another copy in the LSC or SSC regions, whereas 20 are copied within the IR regions. In addition, 18 genes in the *L. ledebourii* chloroplast genome contained introns, including 12 proteins encoding genes (*rps12*, *rps16*, *rpl2*, *rpl16*, *petB*, *petD*, *ndhA*, *ndhB*, *clpP*, *ycf3*, *rpoC1*, and *atpF*) and 6 tRNAs (*trnG-UCC*, *trnL-UAA*, *trnK-UUU*, *trnV-UAC*, *trnA-UGC*, and *trnI-GAU*), of which 18 genes with a single intron and two genes (*clpP* and *ycf3*) with two introns. *infA* was interpreted as a pseudogene. Trans-splicing was observed in the *rps12* gene, with the 5' exon positioned in the LSC region and the intron and 3' exon positioned in the IR regions (Fig. 1, Table 1).

Genome comparison: boundaries regions and divergence hotspot. The *Lilium* cp genomes size differed between 151,655 bp in *L. bakerianum* and 153,235 bp in *L. fargesii*. The expansion and contraction variability in IR/SC junction regions, which were typical phenomena in the plant species evolutionary scrutiny, were evaluated by comparing the border regions and adjacent genes of *Lilium* chloroplast genomes. In this study, the

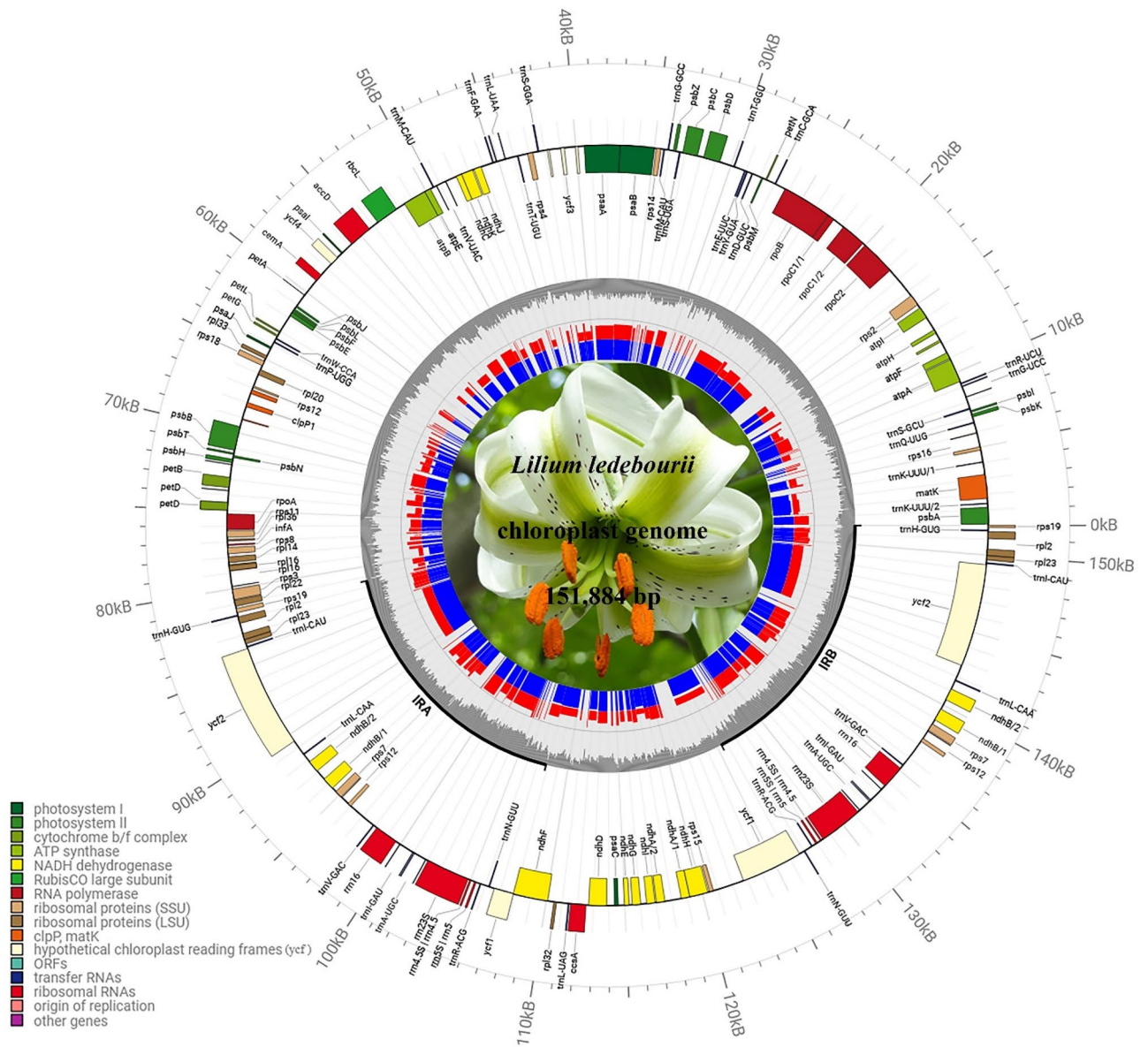


Figure 1. The chloroplast genome map of *L. ledebourii*. Transcriptional directions are represented on the circle's inside (clockwise) and outside (counterclockwise). Genes are color-coded according to their functional groups.

chloroplast genomes of *Lilium* species demonstrated slightly visible junction variation in the IRA/LSC and IRB/SSC boundaries, despite the gene number and gene content being conserved (Fig. 2).

In all *Lilium* cp genomes, IR regions were found to have nearly the same size (26,382 bp to 26,990 bp). *L. fargesii* had the largest IR region expansion, which ended at the *rps19* gene. In all *Lilium* cp genomes, the *ycf1* gene was placed in the IR/SSC boundary regions, leading to an incomplete gene duplication within IRs, and the IR/LSC junction site was placed in the *rps19* gene. In most *Lilium* cp genomes, the *rps19* gene was observed in the IRb location around 137 bp to 142 bp far from the JLB boundary, except for *L. distichum* (6 bp); *L. gongshanense* and *L. nepalense* (22); *L. henricii*, *L. meleagrimum*, and *L. rosthornii* (23 bp); and *L. amoenum* and *L. primulinum* (31 bp): it was found significantly lower than the other species. The *ndhF* gene in most cp genomes (39/48) was positioned inside the SSC location. However, in nine species, including *L. bakerianum* 25 bp, *L. fargesii* 55 bp, *L. gongshanense* 33 bp, *L. lophophorum* 19 bp, *L. pardalinum* 44 bp, *L. philadelphicum* 22 bp, *L. regale* 4 bp, *L. superbum* 45 bp, and *L. washingtonianum* 45 bp, the *ndhF* gene was discovered somewhat extended in the IRb (Fig. 2).

In the *Lilium* species, the IR region was more conserved than LSC and SSC regions. Synteny results revealed that the *Lilium* species have a high degree of sequence identity and collinearity at the cp genome-wide scale, especially in the IR region (Fig. S1). We also conducted genome-wide analysis via sliding window assessment to detect hotspot regions in the *Lilium* cp genomes. Nucleotide diversity (π) was averaged at 0.00504, ranging from 0 to 0.01913. The variability of *rpl32-trnL-ccsA*, *petD-rpoA*, *ycf1*, *psbI-trnS-trnG*, *rps15-ycf1*, *trnR*, *trnT-trnL*, and *trnP-psa]-rpl33* were higher among the 48 *Lilium* cp genomes. The divergence was more prominent in the SC regions than in the IRs regions, which displayed a higher nucleotide variability compared to IR regions (Fig. 3).

Category	Gene group	Gene name
Photosynthesis pathways	ATP synthase	<i>atpA, atpB, atpE, atpF*, atpH, atpI</i>
	NADH-dehydrogenase	<i>ndhA*, ndhB*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK,</i>
	Cytochrome b/f complex	<i>petA, petB*, petD*, petG, petL, petN</i>
	Photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>
	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Rubisco	<i>rbcl</i>
Transcription and translation related genes	DNA-dependent RNA polymerase	<i>rpoA, rpoB, rpoC1*, rpoC2</i>
Ribosomal proteins	Large subunit of ribosomal proteins	<i>rpl2*, rpl14, rpl16*, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i>
	Small subunit of ribosomal proteins	<i>rps2, rps3, rps4, rps7, rps8, rps11, rps12**, rps14, rps15, rps16*, rps18, rps19</i>
RNA genes	Ribosomal RNA	<i>rrn4.5 (×2), rrn5 (×2), rrn16 (×2), rrn23 (×2)</i>
	Transfer RNA	<i>trnA-Ugc* (×2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnM-CAU, trnG-GCC, trnG-UCC*, trnH-GUG (×2), trnI-CAU (×2), trnI-GAU* (×2), trnK-UUU*, trnL-CAA (×2), trnL-UAA*, trnL-UAG, trnM-CAU, trnN-GUU (×2), trnP-UGG, trnQ-UUG, trnR-ACG (×2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC (×2), trnV-UAC*, trnW-CCA, trnY-GUA</i>
Other genes	Maturase K	<i>matK</i>
	Subunit of acetyl-CoA carboxylase	<i>accD</i>
	C-type cytochrome synthesis gene	<i>ccsA</i>
	Envelope membrane protein	<i>cemA</i>
	ATP-dependent protease subunit P	<i>clpP**</i>
	Translational initiation factor	<i>infA^ψ</i>
	Conserved hypothetical open reading frames	<i>ycf1, ycf2, ycf3**, ycf4</i>

Table 1. Gene content and functional classification of *L. ledebourii* chloroplast genome. *Gene with one intron; **gene with two introns; (×2) duplicated gene; ^ψtrans-spliced gene; ^ψ pseudogene.

SSRs and complex repeat analysis. In population genetic studies, the number and position of repeated DNA motifs (with 1–6 nucleotides) have been routinely employed for the detection of polymorphisms in cp genomes⁴⁴. We discovered SSRs in the cp genomes of SCh and 47 closely related species. 64 SSRs were found in the SCh cp genome, mostly (60.94%) made up of mononucleotide repeats. In addition, only one pentanucleotide SSR pattern was observed in SCh (Fig. 4A). As shown in Fig. 4C, among the 48 *Lilium* cp genomes, SSRs ranged from 53 (*L. superbum*) to 81 (*L. pardanthinum*). In total, 3234 microsatellites were detected in 48 cp genomes of *Lilium* (Fig. 4C), with mononucleotide SSRs (57.48%) being the most common, whereas di-, tri-, tetra-, penta-, and hexa nucleotide SSRs accounted for 17.56%, 7.58%, 14.78%, 2.35%, and 0.25% of all SSRs, respectively (Fig. 4B). The number of mono-nucleotide repeats in the 48 *Lilium* cp genomes varied from 25 (*L. distichum*) to 50 (*L. fargesii*). Hexanucleotide repeats were only observed in the cp genome of *L. henricii* (AACTAG/AGTTCT), *L. leichtlinii* (AAATAT/ATATTT and ACTCAT/AGTATG), *L. sp_KHK-2014* (ACGTAT/ACGTAT), *L. tsingtauense* (ACGTAT/ACGTAT), *L. meleagrimum* (AACTAG/AGTTCT), *L. pardalinum* (AATAGT/ACTATT), and *L. sargentiae* (AAATTC/AATTTG). Overall, varied SSR motifs were found in *Lilium* cp genomes at different frequencies. This research distinguished the presence and SSR types of *Lilium* species, which might be fruitful for molecular marker investigations and population genetics of *Lilium*, especially for *L. ledebourii* (Table S1, Fig. S2).

Complicated repeat sequences play a role in the recombination and variation of chloroplast genomes⁴⁵. The SCh chloroplast genome contains 32 complex repeats, including five tandem, ten dispersed, and 17 palindromic repeats (Table 2). These repeats were at least 24 bp in length, with the longest being 53 bp. Furthermore, it was discovered that the final quantity of complex repeats in the SCh genome was around 25% lower than the average number of repeats in *Lilium* genomes, with a decrease of 18%, 32%, and 22% for tandem, dispersed, and palindromic repeats, respectively (Fig. 4D). In total, 29–55 long repeat sequences were discovered in each *Lilium* cp genome, including 9–23 dispersed repeats, 14–30 palindromic repeats, and 3–10 tandem repeats (Fig. 4D). With the exception of *L. longiflorum*, which contained a repetition of 162 bp, repeat sizes varied from 30 to 87 bp in dispersed, 30 to 61 bp in palindromic, and from 15 to 85 bp in tandem. In other words, out of the 48 species, palindromic repeat was the most common type, and the total number of repeats ranged from 29 to 55, with 60.86% of these repeats being between 30 and 40 in length (Fig. 4E).

Codon usage bias analysis. Due to the widespread occurrence of synonymous codon bias in organisms, recognizing codon preference might play a significant role in the evolution by clarifying the selection pressure and improving translation efficiency by utilizing major codons^{46,47}. Totally, 21,989 codons were detected in the SCh protein-coding genes. A- and U-ending are seen to be more prevalent than G and C-ending ones. Among SCh amino acids, the highest and lowest frequencies were related to leucine (Leu = 2268) and cysteine (Cys = 255), respectively. In SCh, 30 codons showed more bias (RSCU > 1), and 31 codons displayed bias: RSCU < 1. In addition, there was no bias (RSCU = 1) in the frequency of start codons AUG (methionine), UGG (tryptophan), and AUA (isoleucine) (Table 3).

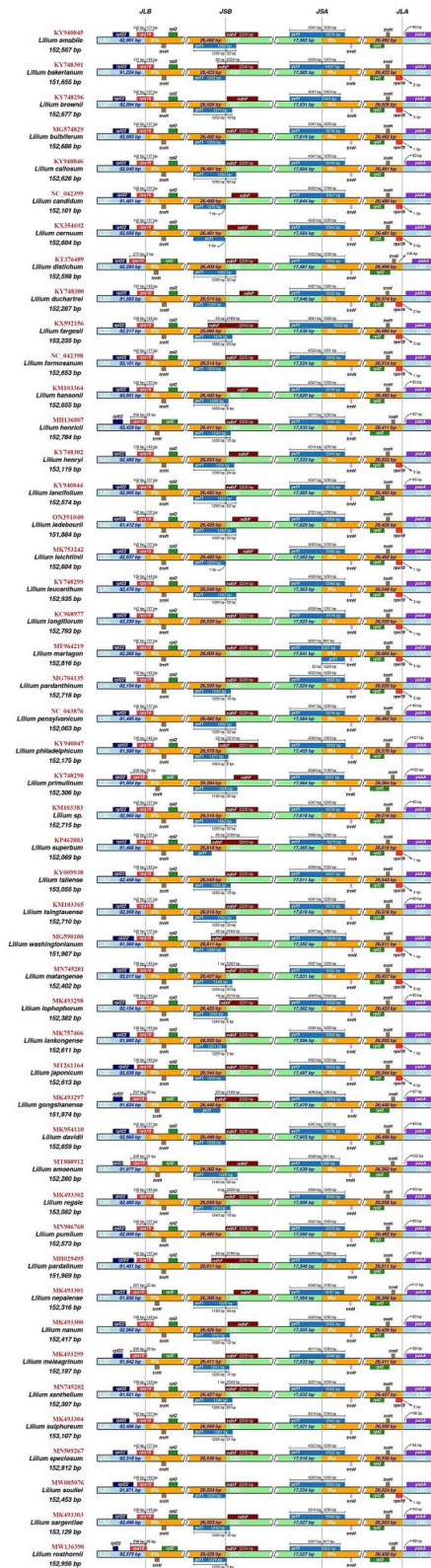


Figure 2. Comparison of the junction positions of LSC, SSC, and IR regions among the among 48 Lilium cp genomes. The red identifiers represent the GenBank accession number of each species.

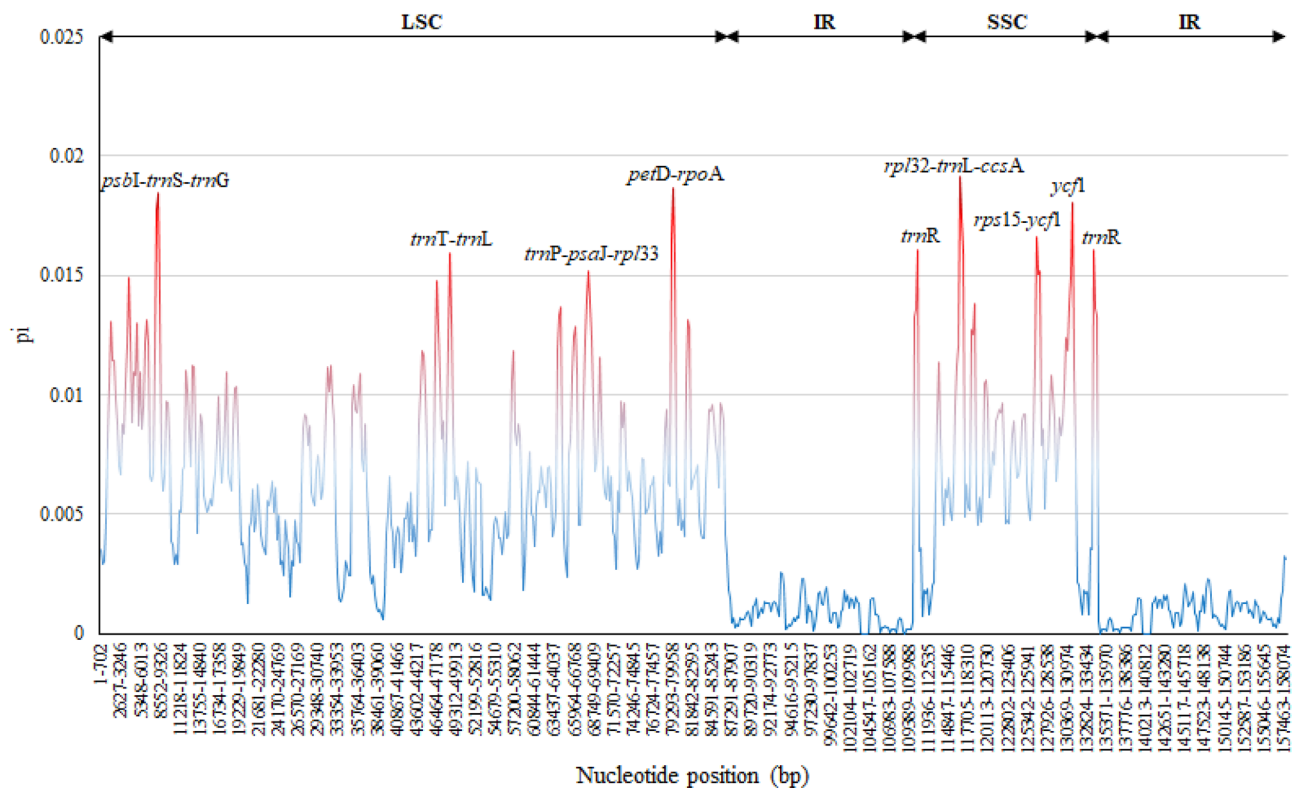


Figure 3. Sliding window analysis of 48 *Lilium* cp genomes (window length: 600 bp—step size: 200 bp). The X-axis and Y-axis represents the position of a window and nucleotide diversity (Pi) of each window, respectively.

Comparing the protein-coding genes in 48 *Lilium* cp genomes, we found that each species was composed of 20,691–22,781 triplet codons in protein-coding genes. Leucine (10.18–10.34%) was the most abundant among encoded amino acids in all of the species studied, whereas cysteine (1.13–1.24%) was the least abundant (Table S2, Fig. 5). Among the 20 amino acids, the lowest and highest RSCU values were recorded for Tyr-UAC, encoding the tyrosine, and Leu-CUU, encoding the leucine, respectively. Codon usage in the *Lilium* cp genomes was biased towards A- and U-ended codons, according to RSCU values (RSCU > 1). In addition, the pattern of codon usage bias in the *Lilium* species was investigated. Figure S3 shows the values of the Codon adaptation index (CAI), Codon bias index (CBI), Frequency of optimal codons (FOP), Effective number of codons (NC), and GC3s for 48 *Lilium* chloroplast genomes. We observed that five parameters associated with codon usage bias are very similar across *Lilium* species (Fig. S3).

Selection pressure on *Lilium* cp genomes and adaptation evolution. In this study, the Ka/Ks ratio was computed for the 78 protein-coding genes shared by all 48 cp genomes (Table S3). Ks = 0 may lead to swelling of Ka/Ks ratios or misidentification of genes with powerful positive selection because of high, unlimited, or unspecified Ka/Ks ratio. To put down this difficulty, we eliminated the comparisons with Ks = 0 from all analyses. The average Ka/Ks value of the 78 protein-coding genes examined among 48 cp genomes was 0.2472. Among the genes, *ndhB*, *psbJ*, *psbZ*, and *ycf2* exhibit zero synonymous substitution (Ks = 0), revealing divergence of species or genetic restriction. The *rps7*, *atpH*, *petN*, *psaI*, *psaJ*, *psbF*, *psbL*, *psbT*, and *psaC* were the highly conserved genes, with an average Ka/Ks ratio of 0, indicating extremely purifying selection pressure (Table S3). We found that *accD*, *rpl16*, and *rpl36* with Ka/Ks average of 1.268, 1.052, and 1.200, respectively, have been subjected to positive selection in the *Lilium* cp genomes. The *matK*, *petB*, *petD*, *rps4*, *rps12*, and *ycf1* had average Ka/Ks ratios in the 0.5 to 1 range, reflecting calm selection. The Ka/Ks average for the rest genes was recorded less than 0.49, reflecting that about 72% of genes (56/78) in the *Lilium* cp genomes were subject to purifying selection. Although the average Ka/Ks > 1 was recorded only for the *accD*, *rpl16*, and *rpl36*, for 26 genes, Ka/Ks > 1 was observed in at least one pairwise comparison (from 1128 pairwise comparisons). The gene *ycf1* possesses 248 positive selective pairwise comparisons, followed by *matK* (144), *ccsA* (52), *rbcl* (37), *ndhI* (31), *clpP* (29), *atpF* (20), *rpoC2* (20), *cemA* (16), *ndhD* (14), *ndhF* (10), *petB* (10), *ndhA* (9), *ndhG* (9), *petD* (9), *ycf4* (8), *rpoA* (7), *ndhJ* (4), *rpl33* (3), *rps14* (3), *ndhH* (2), *petG* (2), *rps4* (2), *ndhC* (1), *rpoB* (1), and *rps2* (1).

Also, another analysis to compare *Lilium* species based on “concatenate all protein-coding genes” was performed and obtained 1128 pairwise comparison outcomes of Ka/Ks values. The pairwise Ks value of zero was seen between *L. matangense* and *L. xanthellum*. Ka/Ks > 1 was only obtained between *L. speciosum* and *L. distichum* with a Ka/Ks ratio of 1.375. The Ka/Ks = 1 was recognized between *L. amabile* and *L. pumilum*, *L. henricii* and *L. meleagrimum*, *L. sulphureum* and *L. henryi*, and *L. lancifolium* and *L. pumilum* (Ka/Ks = 1.062). Ka/Ks = 0 among all comparisons, only was recorded between *L. sp._KHK-2014* and *L. tsingtauense*. Overall, the average pairwise

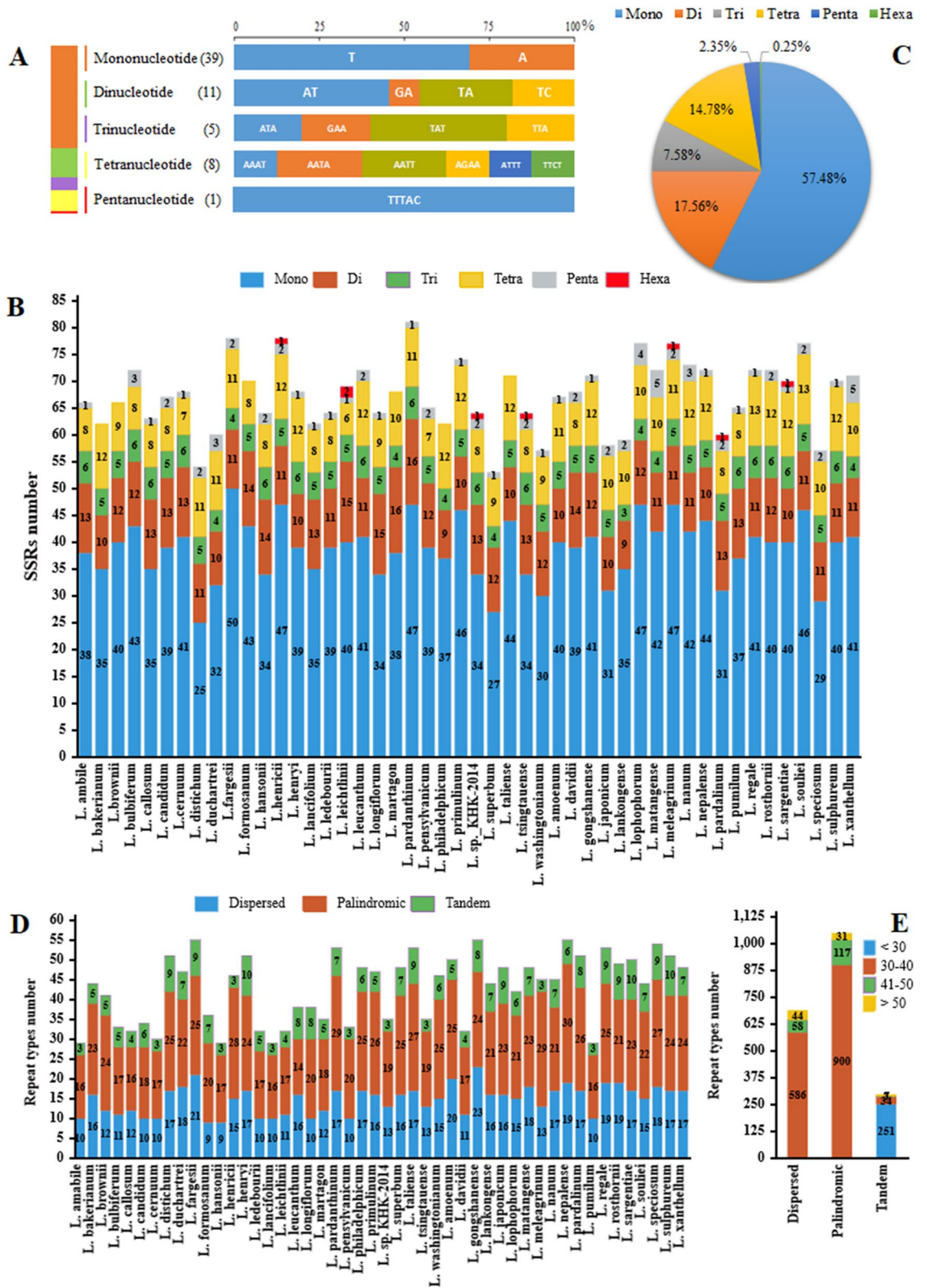


Figure 4. The type and distribution of simple sequence repeats (SSRs) and complex repeat in the 48 *Lilium* cp genomes. (A) Frequency and type of SSRs in the *L. ledebourii* cp genome. (B) The number of SSR types discovered in 48 *Lilium* cp genomes. (C) The percentage of SSRs types in 48 *Lilium* cp genomes. (D) The number and of complex repeats types in 48 *Lilium* cp genomes. (E) Frequency of complex repeats by size.

N	Size (bp)	Start position1	Type	Start position2	E-value
1	53	36,308	D	38,532	5.06E-17
2	43	43,951	P	43,951	1.08E-14
3	38	27,465	P	27,465	8.59E-14
4	34	112,143	P	112,143	2.20E-11
5	40	88,672	D	88,696	3.77E-11
6	40	88,672	P	144,558	3.77E-11
7	40	88,696	P	144,582	3.77E-11
8	40	144,558	D	144,582	3.77E-11
9	39	41,001	D	96,049	1.43E-10
10	39	41,001	P	137,206	1.43E-10
11	41	71,433	P	71,433	3.86E-10
12	35	91,474	P	91,474	5.77E-10
13	35	91,474	D	141,785	5.77E-10
14	35	141,785	P	141,785	5.77E-10
15	34	53,789	P	53,789	1.11E-07
16	35	33,896	P	33,896	9.71E-07
17	34	75,092	D	75,118	3.55E-06
18	31	7496	P	42,412	5.89E-06
19	33	36,334	D	38,558	1.30E-05
20	30	3602	P	68,653	2.20E-05
21	30	5488	P	5488	2.20E-05
22	30	120,861	P	120,896	2.20E-05
23	32	33,306	P	42,412	4.71E-05
24	31	7496	D	33,307	1.71E-04
25	30	9028	D	34,112	6.17E-04
26	30	29,142	D	29,414	6.17E-04
27	30	110,845	P	110,873	6.17E-04

Table 2. Dispersed and palindromic repeats by positions in the cp genome of *L. ledebourii*.

Ka/Ks value was 0.4839 (Fig. 6), indicating that at the whole cp protein level, *Lilium* species were subjected to a purifying selection.

Phylogenetic analysis. Both analyses, which used the complete chloroplast genome (CCGs) and protein coding genes (CDSs), divided 47 *Lilium* species into two main groups. Although the members of both main groups were the same in both topologies, there were slight differences between them (Figs. 7, S4). Because of the high synteny among *Lilium* cp genomes, this study concentrated on phylogenetic analysis employing whole cp genome sequences to inspect relationships across the 47 *Lilium* species. Maximum likelihood was employed with two species serving as outgroups. According to the topology, the majority of nodes were highly supported. In all, 34 of the 45 nodes acquired a maximally supported (value $\geq 99\%$) value bootstrap. According to the CCG topology, the 47 *Lilium* species were divided into two main groups consisting of 11 clades (Fig. 7).

Group 1: included 17 species, which were placed in four clades and four self-sufficient lineages. Clade I consisted of six Sinomartagon species including *L. callosum*, *L. amabile*, *L. cernuum*, *L. lancifolium*, *L. pensylvanicum*, and *L. primulinum*. Clade II is composed of two Martagon species (*L. tsingtauense* and *L. hansonii*). Clade III consisted of two Leucolirion (*L. formosanum* and *L. longiflorum*), and *L. brownie* belong to Archelirion. *L. candidum* of Liriotypus and *L. ledebourii* established a monophyletic clade IV. *L. davidii* was the sister of Clade II. *L. martagon*, *L. leichtlini*, and *L. bulbiferum* had more distant connections with clade I Sinomartagons (Fig. 7).

Group 2: included 30 species, which were placed in seven clades and three self-sufficient lineages. Clade V was composed of four Sinomartagon species including *L. bakerianum*, *L. taliense*, *L. primulinum*, and *L. nepalense*. Clade VI consisted of two Sinomartagon (*L. henricii* and *L. souliei*), and three species (*L. pardanthinum*, *L. gongshanense*, and *L. meleagrimum*), which have recently been transferred from the genus *Nemacaris* to the genus *Lilium*. *L. speciosum*, *L. japonicum*, and *L. distichum* formed clade VII. Five species of Leucolirion including *L. regale*, *L. leucanthum*, *L. sulphureum*, *L. sargentiae*, and *L. henryi*, created clade VIII. Clade IX accommodated *L. duchartrei* and *L. lankongense*. Three species of Loforum (*L. fargesii*, *L. lophophorum*, and *L. matangense*), along with *L. nanum* and *L. xanthellum*, were included in clade X. Three species (*L. superbum* and *L. washingtonianum*, and *L. pardalinum*) belonging to Pseudolirium formed clade XI. *L. amoenum*, *L. rosthornii*, and *L. philadelphicum* were sisters to the clade V, clade VIII, and clade X, respectively (Fig. 7).

Codon	AA	ObsFreq	RSCU	Codon	AA	ObsFreq	RSCU
UAA	^a	68	1.468	AUG	M	552	1
UAG	^a	41	0.885	AAC	N	216	0.422
UGA	^a	30	0.647	AAU	N	808	1.578
GCA	A	345	1.151	CCA	P	254	1.121
GCC	A	186	0.621	CCC	P	184	0.812
GCG	A	125	0.417	CCG	P	116	0.512
GCU	A	543	1.812	CCU	P	352	1.554
UGC	C	62	0.486	CAA	Q	573	1.504
UGU	C	193	1.514	CAG	Q	189	0.496
GAC	D	175	0.411	AGA	R	401	1.519
GAU	D	676	1.589	AGG	R	127	0.481
GAA	E	859	1.511	CGA	R	298	1.53
GAG	E	278	0.489	CGC	R	81	0.416
UUC	F	435	0.691	CGG	R	109	0.56
UUU	F	824	1.309	CGU	R	291	1.494
GGA	G	588	1.581	AGC	S	86	0.383
GGC	G	171	0.46	AGU	S	363	1.617
GGG	G	253	0.68	UCA	S	352	1.149
GGU	G	476	1.28	UCC	S	259	0.846
CAC	H	108	0.408	UCG	S	144	0.47
CAU	H	421	1.592	UCU	S	470	1.535
AUA	I	647	1	ACA	T	353	1.253
AUC	I	365	0.564	ACC	T	205	0.728
AUU	I	929	1.436	ACG	T	115	0.408
AAA	K	837	1.499	ACU	T	454	1.611
AAG	K	280	0.501	GUA	V	456	1.482
CUA	L	289	1.115	GUC	V	161	0.523
CUC	L	160	0.617	GUG	V	167	0.543
CUG	L	135	0.521	GUU	V	447	1.452
CUU	L	453	1.747	UGG	W	399	1
UUA	L	782	1.271	UAC	Y	162	0.393
UUG	L	449	0.729	UAU	Y	662	1.607

Table 3. The Relative synonymous codon usage (RSCU) of *L. ledebourii* protein-coding genes. ^aStop codon.

Discussion

The conserved characteristics of gene content and organization, as well as the GC content of the SCh cp genome, were found to be similar to the variability in other species^{17,18}. Furthermore, trans-splicing was observed in the *rps12* gene, which is also seen in other species⁴⁸. The *Lilium* cp genomes length differed between 151,655 bp in *L. bakerianum* and 153,235 bp in *L. fargesii*. It was suggested that one of the primary causes for the change in the cp genomes size is that the IR region shrinks, expands, or losing⁴⁹.

We surveyed the expansion and contraction variability in IR/SC junction regions. The boundary of LSC/IRb is stable, while slightly visible junction variation can be seen in the IRa/LSC and IRb/SSC boundaries. The occurrence of contraction and expansion of IR regions during evolution is a relatively common happening, which has been employed as evolutionary loci for phylogenetic studies^{50,51}. Expansion of IR/LSC junctions to *rps19* has been observed in other Liliaceae species such as *Amana*, and this event appears to be a Liliaceae ancestral symplesiomorphy⁵². The contraction/expansion of IR regions in Liliaceae has resulted in the formation of *ycf1* and *rps19* at the boundaries across SC and IR regions, with varying lengths, as demonstrated in *Fritillaria*⁵³ and *Cardiocrinum*⁵⁴. Given the unison of our findings with those about other plants^{52–54}, expansion and contraction of IR regions may be a significant mechanism for different lengths of 48 *Lilium* cp genomes.

Analysis of nucleotide diversity and cp repeats can be used to recognize molecular markers, rebuild evolutionary connections, and delve into population genetics^{55,56}. In this study, a total of 3234 microsatellites were discovered in 48 cp genomes of *Lilium*. A/T repeats were the most common type of SSR found. The abundance of this SSR type is consistent with the majority of other cp genomes explored thus far⁵⁷. Additionally, complex repeats in the 48 *Lilium* were found, which could be substantial genomic reconfiguration hotspots^{58,59}. The number and size of tandem, dispersed, and palindromic repeats were nearly identical in the cp genomes of relevant species such as *Fritillaria*⁶⁰. Especially, the incidence of large repeats in the chloroplast, such as the 162 bp tandem repeat of *L. longiflorum* from our results, has probably been linked to an unstable genomic structure because of improper rearrangements⁶¹.

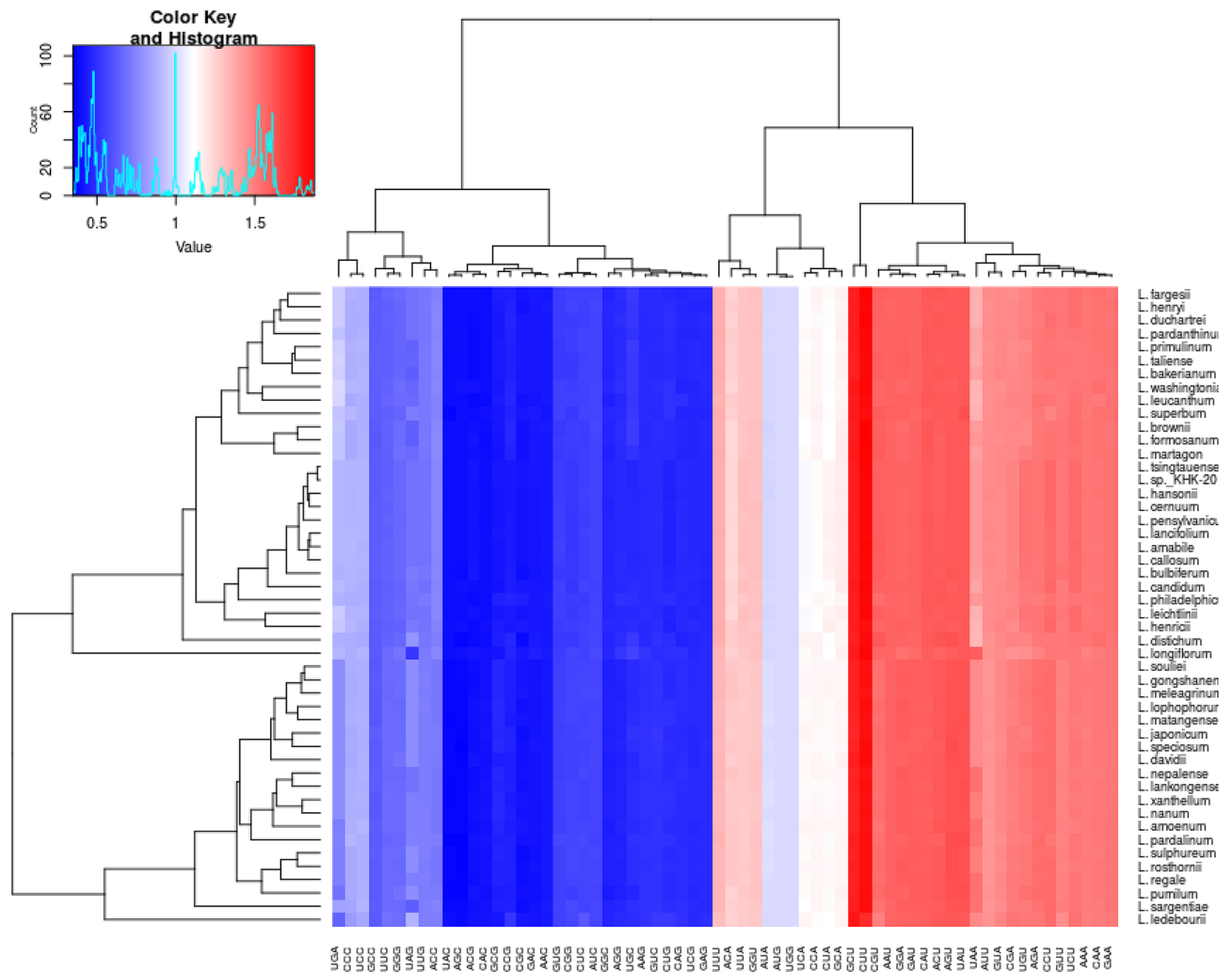


Figure 5. Codon distribution of protein-coding genes among *Lilium* cp genomes. Color code: Red denotes a higher RSCU and blue denotes a lower RSCU.

Here, a higher divergence of the SC regions and a lower divergence of the IR were discovered, suggesting that the IR is more conservative than in other regions, with the same characteristics of most angiosperms⁶². This occurrence is caused by copy correction of IRs and the removal of harmful mutations via gene conversion⁶³. *rpl32-trnL-ccsA*⁶⁴, *trnP-psa]-rpl33*, *petD-rpoA*, *ycf1*¹⁸, *psbI-trnS-trnG*⁶⁵, and *rps15-ycf1*, *trnT-trnL*⁶⁶ have previously been identified as high variability regions in various species. The phylogeny of genus *Lilium* will be clarified with the help of these regions, which are expected to be very helpful in the future.

The rate of K_a (non-synonymous) to K_s (synonymous) nucleotide substitution is commonly employed as a powerful tool for the clarification of the evolution of protein coding genes and also species adaptive developments^{67,68}. The K_a/K_s ratio determines the gene divergence grade and whether selection pressure is positive ($K_a/K_s > 1$), purifying ($K_a/K_s < 1$, particularly if it is less than 0.5), or neutral ($K_a/K_s = 1$)⁶⁷.

In the present study, the $K_a/K_s > 1$ was recorded for *accD*, *rpl16*, and *rpl36*, implying that these genes could be important in the adaptive evolution⁷¹. Positive selection of *accD* was signed by the important role of the gene in stress tolerance and resistance, insect predation, and pathogens⁶⁹. Positive selection in *accD* has been observed in *Ipomoea*⁴¹ and *Stauntonia*⁴³. Other positively selected genes were *rpl16* and *rpl36*, which are responsible for encoding the ribosomal protein, which has been evidenced to be necessary for the development of chloroplast ribosomes in plants⁷⁰. Previous studies have reported positive selection for *rpl16* in *Lonicera*⁷¹ and *rpl36* in *Aquilegia*⁴².

However, we discovered that some genes were positively selected in at least one pairwise comparison, suggesting these genes were potentially subject to positive pressure for selection among *Lilium* species. *ycf1* possesses 248 positive selective pairwise comparisons, followed by *matK* (144), *ccsA* (52), *rbcl* (37), *ndhI* (31), *clpP* (29), *atpF* (20), *rpoC2* (20), *cemA* (16), *ndhD* (14), *ndhF*(10), *petB* (10), *ndhA* (9), *ndhG* (9), *petD* (9), *ycf4* (8), *rpoA* (7), *ndhJ* (4), *rpl33* (3), *rps14* (3), *ndhH* (2), *petG* (2), *rps4* (2), *ndhC* (1), *rpoB* (1), and *rps2* (1). Of these, *MatK* has previously been found under positive selection in over 30 different taxonomic groups⁷². NADH-dehydrogenase gene group (*ndh*) were fundamental in the use of light energy and the electron transfer chain to produce ATP, significant components for photosynthesis^{73,74}. As a consequence, these genes, as important components involved

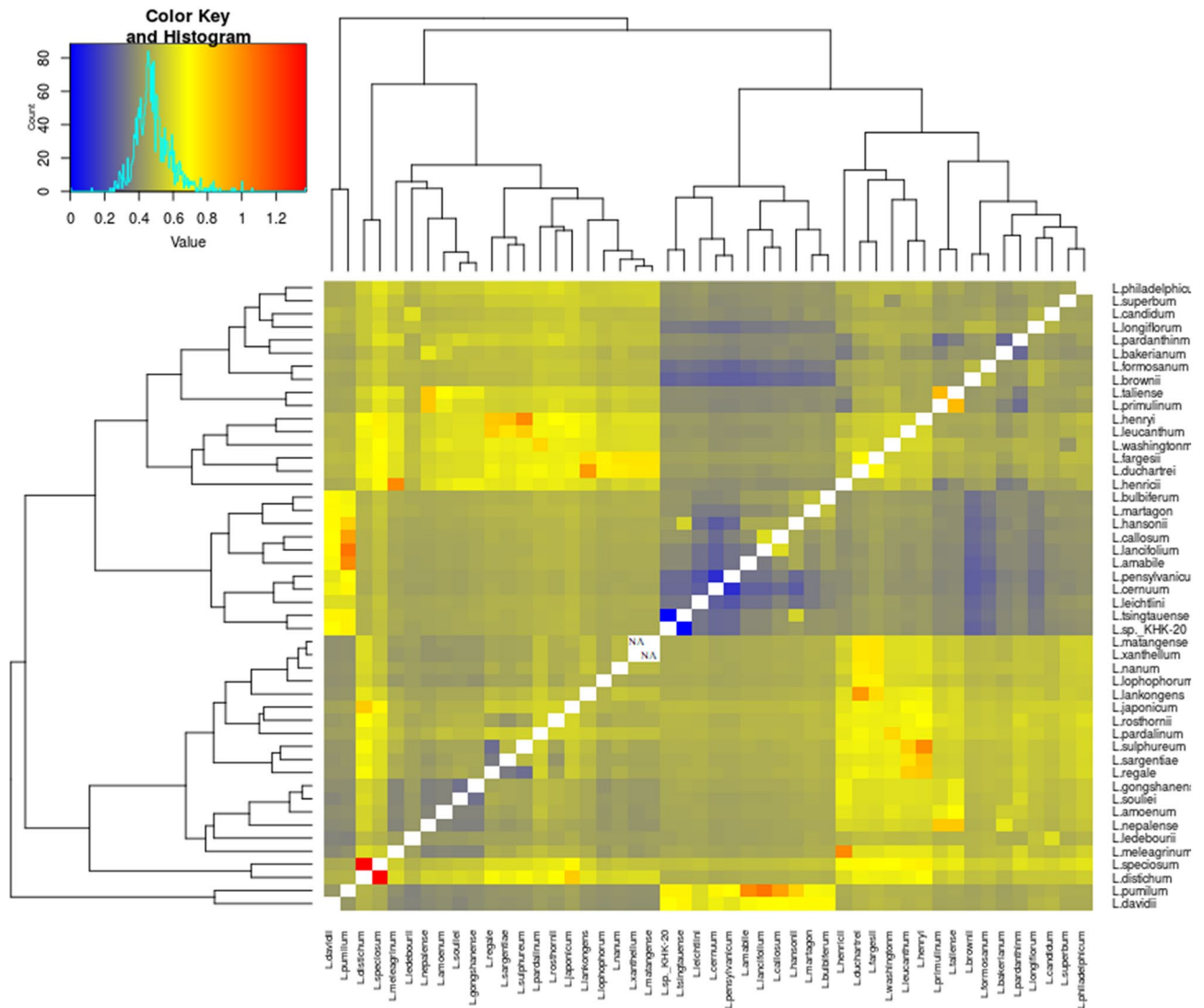


Figure 6. Ka/Ks ratios between *Lilium* cp genome pairs. In the multigenic nucleotide alignment, the heatmap depicts pairwise Ka/Ks ratios between each concatenated single-copy CDs sequence.

in plant growth, may have evolved as a result of more common substitutions in order to adapt to various environmental conditions⁴³. We discovered positive selection on the *ycf1* gene in 248 pairwise comparisons. The *ycf1* is huge open reading frame, which encodes protein products for many amino acids in higher plant. Moreover, the necessity of the *ycf1* gene for cell survival has been proven by knockout studies⁷⁵. *clpP*, which encodes the (ATP-dependent) *clp* protease, is thought to play a role in chloroplast protein transformation and may be necessary for shoot development in the presence of the degradation of *clpP*-mediated protein^{76,77}. Another positively selected gene in our study is the rubisco large-chain gene (*rbcL*). In many higher plants, *rbcL* positive selection has been made⁷⁸. *atpF* is involved in the encoding of the H⁺-ATP subunits, which is necessary for some photosynthetic processes⁷⁹. Positive selection has been noticed to evolve *rpo* genes, which encode proteins role-playing in the modification of transcription and post-transcriptional modification⁸⁰. Due to the cooperation of *cemA* with nuclear genes⁸¹, *cemA* might evolve relatively quickly in species⁸². Substitution of amino-acid, indel presence and prematurity in stop codon could lead to a positive selection of *ccsA*⁸³.

Selected positive genes may have had crucial roles in the adaptation of *Lilium* to different environments. Gao et al.⁸⁴ documented the adaptation of chloroplast genes to various ecological environments of solar preferences. Moreover, more undiscovered selective compulsions may be involved in the increasing of the Ka/Ks ratio, leading to species divergence⁸⁵. The Ka/Ks ratios in the majority of genes shared by *Lilium* chloroplast genomes and among pairwise comparisons of species employing all protein-coding genes were less than 1, proposing purifying selection. Similar findings were detailed for *Gentiana* species⁸⁶. This lower rate of Ka/Ks can be the result of the fact that most of the species are probably to undergo disadvantageous nonsynonymous substitutions and purification selections, and the selective restriction on nonsynonymous substitutions is stronger than synonymous substitutions^{87,88}. In short, positive selection in some genes likely enriches the *Lilium* variety and adaptability.

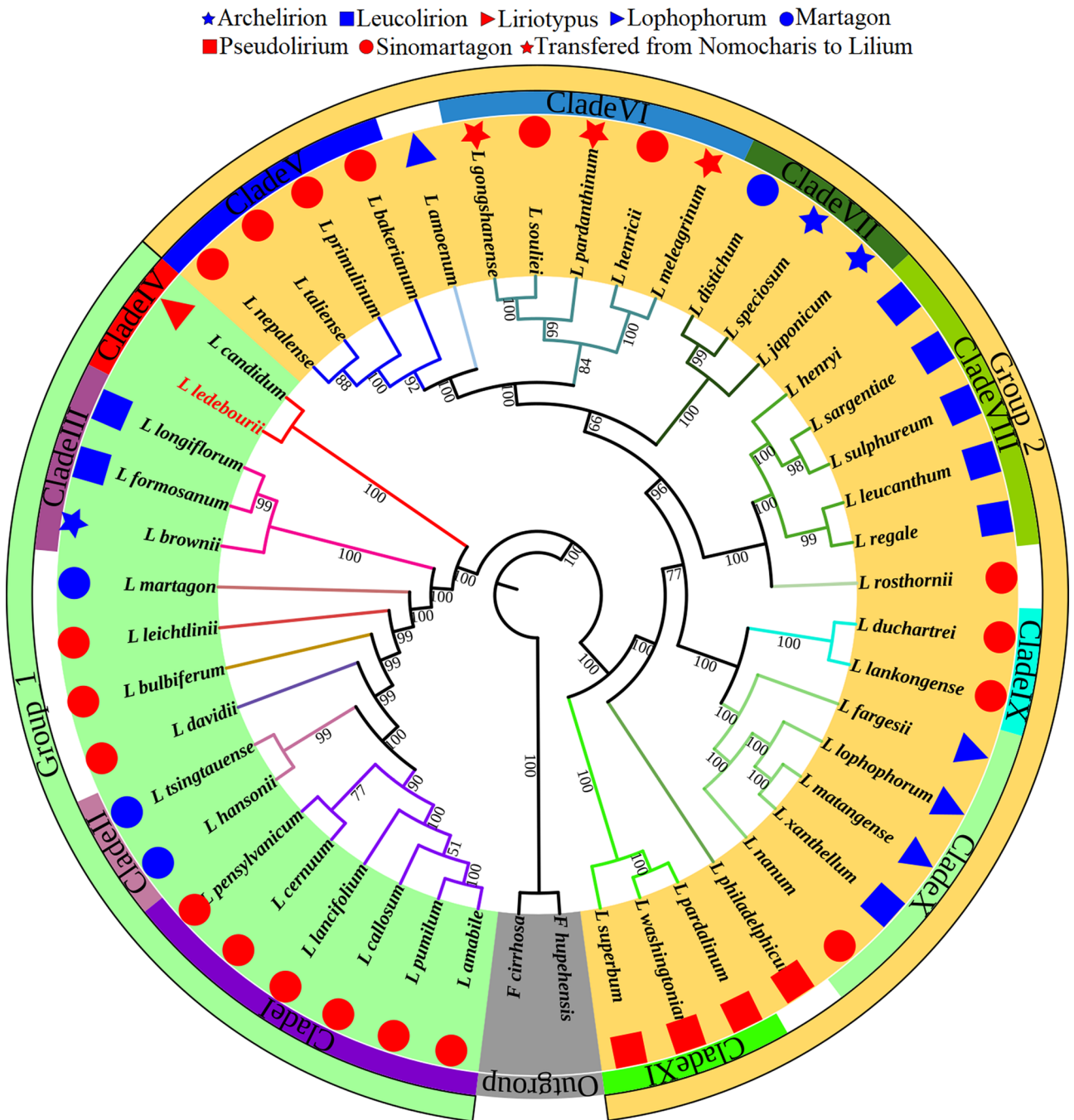


Figure 7. The phylogenetic relationships of *Lilium* species employing whole cp genome sequences. *Fritillaria hupehensis* and *Fritillaria cirrhosa* were applied as outgroups. Phylogenetic tree were constructed by Maximum likelihood (ML). The ML bootstrap values are represented by the numbers above the branches.

At the whole cp protein scale, *Lilium* species were subjected to a purifying selection. Sunlight UV radiation damages and rearranges DNA^{89,90}, and higher temperatures speed up metabolism⁹¹, all contributing to an increase the mutation rates. Consequently, purifying selection, as one of the most common types of natural selection, constantly helps in the elimination of disadvantageous mutations in populations. Purifying selection would thus be an evolutionary outcome of the preservation of *Lilium* species adaptive habits.

The richer the taxon sampling, the more accurate it is to comment on the Comber’s classification. The present study clustered 47 species of *Lilium*. To date, some of them have not been evaluated at the whole cp genome level. *Lilium* species were distributed into 11 clades divided into two main groups. The position of the species in the present topology is consistent with the classification by Kim et al.¹².

Our samples did not support the monophyly of Comber’s sections¹¹. According to our results, *L. martagon* was placed farther away from the Martagon species and was closer to *L. leichtlinii*, *L. bulbiferum*, and *L. davidii*, which agrees with Gong et al.¹⁶ based on nrITS. We observed *L. brownii* of Comber’s Archelirion far from the

other two species of Archelirion) *L. japonicum* and *L. speciosum* (and close to the two species of Leucolirion (*L. formosanum* and *L. longiflorum*). Similarly, Li et al.⁹ classified *L. brownii* alongside *L. formosanum* and *L. longiflorum* in a genus-level study of the Liliaceae family's evolution. As our results warn, and with the help of ITS-dataset classification^{92,93}, now with the approval of cp genome-based classification, *L. brownii* can be moved from the Archelirion to the Leucolirion.

Based on the morphology, *L. duchartrei*, *L. lankongense*, *L. nanum*, and *L. rosthornii* belong to the Sinomartagon section¹¹. However, based on our cp genome-scale topology, these species are further away from the Sinomartagons and are closer to the Lophophorum species. Studies have shown that *L. nanum* and Lophophorum have similar karyotypes⁹⁴. Additionally, according to the ITS regions, Du et al.⁹² reported that *L. duchartrei* and *L. lankongense* are in the same clade and are closer to the Lophophorum. We accommodated *L. xanthellum* on Clade X, away from Leucolirion. According to ultrametric chronograms, *L. xanthellum* is closer to *L. lophophorum* and *L. matangense*⁹⁵. Totally, the composition of clade X in our phylogeny (three species of Lophophorum including *L. fargesii*, *L. lophophorum*, and *L. matangense*, along with *L. nanum* of Sinomartagon and *L. xanthellum* of Leucolirion) is in agreement with Du et al.⁹² based on ITS regions.

Although, the monophyletic of the three Martagon species (*L. hansonii*, *L. tsingtauense*, and *L. distichum*) were rejected^{12,92} prior to our study, so far, the position of *L. distichum* has not been clear enough due to restriction of the sampling of Lophophorum and Nomocharis. Based on our topology tree and with the help of a richer sampling, *L. distichum* is further away from its companions in previous studies. Our phylogenetic tree shows *L. distichum* close to two species of Archelirion (*L. japonicum* and *L. speciosum*). Moreover, in terms of flower morphology, Comber's Martagon members have differences from each other. The flowers of *L. distichum* are outfacing, whereas those of *L. tsingtauense* and *L. martagon* are upright and nodding, respectively¹⁹.

As shown in the results (Fig. 7), *L. henricii* and *L. souliei*, two species of Comber's Sinomartagon, are placed next to Nomocharis-Lilium (*L. gongshanense*, *L. meleagrimum*, and *L. pardanthinum*). Gao et al.⁹⁶ based on biogeographic results, showed that *L. henricii* is associated with Nomocharis species. What is more, Gao et al.⁹⁷ by examining 38 *Lilium* species and 7 Nomocharis species using the ITS dataset, showed *L. souliei* inside the Nomocharis clade.

Based on our CCG topology, *L. amoenum* of Lophophorum was sister to *L. bakerianum*, a member of clade V. Zhou et al.⁹⁸, based on fluorescence in situ hybridization, showed the signal pattern of 35S rDNA in *L. amoenum* was the same as in *L. bakerianum*. Moreover, these researchers, by mapping the chromosome pattern for 35S rDNA based on ITS data, showed that these two species are monophyletic.

The phylogenetic position of *L. ledebourii* was very ambiguous due to the scarcity of molecular information. To date, two studies have attempted this. Kim and Kim⁸ involved *L. ledebourii* (with 15 other species) in building the phylogenetic tree. However, this research does not provide a clear picture of the position of this species, due to sampling restriction and the use of only four chloroplast genes (*rbcL*, *matK*, *ndhF*, and *atpB*), which according to the genome-scale results, are not among the most divergent hotspots. Ghanbari et al.⁹⁹, based on the ITS marker, have examined the position of this species and shown that *L. ledebourii* (Damash sample) is far from the *L. candidum*.

Following the resolution of *L. ledebourii*, as one of the study's objectives, and according to Kim et al.¹², who reported that *L. candidum* position remained uncertain, interestingly here, according to cp genome-scale comparisons, *L. ledebourii* and *L. candidum* were monophyletic. *L. ledebourii* is a rare species that has only been seen in Iran and Azerbaijan²⁰. Furthermore, *L. candidum* is thought to have originated in Persia and Syria¹⁰⁰. In total, our findings indicate that a whole cp genome phylogenomic comparison would resolve much controversy and pave the way for *Lilium* phylogeny, especially for *L. ledebourii*.

Conclusions

The whole chloroplast genome of *L. ledebourii* is reported for the first time. The current study using whole chloroplast genomes of *Lilium* revealed structural characteristics, sequence diversity, and enhanced links between species. Meanwhile, certain variation hotspots identified as high variability regions could function as particular DNA barcodes. We provide a comprehensive analysis of selection pressure, and in the whole cp protein scale, *Lilium* species were subjected to a purifying selection. This study covered the restriction of sampling of Lophophorum and Nomocharis as much as possible. For the first time, *L. ledebourii* participated in the classification of genus *Lilium*, and its position was determined. The position of some species, e.g., *L. distichum*, became clearer than before. It is suggested that *L. brownii* can migrate from the Archelirion to the Leucolirion. We believe that the *Lilium* species have been classified with more excellent resolution than in earlier studies, which will be helpful in the understanding of the evolution of *Lilium* species. The genetic resources provided here will aid future studies in species identification, population genetics, and *Lilium* conservation.

Materials and methods

Sample collection and DNA extraction. Fresh leaves of *L. ledebourii* were sampled from Damash village and frozen in liquid nitrogen. The leaf samples were gathered in compliance with national and international legislation and guidelines. It was certified by the herbarium of the Faculty of Agricultural Science and Engineering, University of Tehran, and a validated voucher specimen was deposited at the Department of Horticultural Science, with voucher specimen number 6594. The total genomic DNA was isolated from leaves utilizing a DNeasy plant DNA extraction kit (Qiagen, USA) and the manufacturer's guidelines. DNA integrity was evaluated applying 1% agarose gel, and DNA quantification was assessed employing a NanoDrop spectrophotometer. Extracted DNA was stored at -80°C .

Chloroplast DNA sequencing and genome assembly. SMRT library with a 15–20 Kb insert size was sequenced applying the PacBio RS II platform in Duke Center for Genomics and Computational Biology, USA. To extract potential chloroplast sequences, the PacBio data were mapped to the reference *L. hansonii* cp genome (KM103364) data using BLASR¹⁰¹. Error correction was performed on SMRT reads by sprai pipeline¹⁰². The corrected reads were assembled employing Perl-based pipeline¹⁰³. Furthermore, overlapping ends were checked by “check_circularity.pl” script.

Chloroplast genome annotation. GeSeq¹⁰⁴, with NCBI RefSeq for *Lilium* as the reference dataset, was utilized for the annotation of protein-coding, ribosome RNA (rRNA), and transfer RNA (tRNA) genes. Using BLAST against cp genes, sequence coordinates of all annotated genes were checked and manually edited. The tRNAscan-SE version 2.0 was used to double-check the tRNA genes¹⁰⁵. A circular physical map of the chloroplast genome was illustrated using OrganellarGenomeDRAW (OGDRAW) toolkit¹⁰⁶ and Chloe (<https://github.com/ian-small/chloe>).

Repeat sequence analysis. The MicroSatellite identification tool (MISA) was applied to screen simple sequence repeats (SSRs) in 48 cp genomes with a threshold of 10 for mononucleotide simple sequence repeats (SSRs), 5 for di-, 4 for tri-, and 3 for tetra-, penta-, and hexa nucleotide¹⁰⁷. Tandem repeats were also discovered using default parameters by Tandem Repeats Finder v4.09¹⁰⁸ using default parameters. Moreover, Vmatch V2.3.1¹⁰⁹ was used to identify palindromic repeats (≥ 20 bp) and dispersed (≥ 30 bp).

Chloroplast genome comparison. In order to discover the *Lilium* divergence regions, the distance among adjoining genes and junctions of *L. ledebourii* SSC, LSC, and IRs regions, were compared to the other Lily cp genomes species. For each particularized codon, the ratio of usage frequency was obtained as the Relative Synonymous Codon Usage (RSCU) value using DAMBE V6 for 48 *Lilium* species¹¹⁰. We mapped the results of codon preference via the R program. To evaluate the degree of codons bias, the CodonW V1.4.4 (<http://codonw.sourceforge.net>) calculated the values of the codon bias index (CBI), the codon adaptation index (CAI), the frequency of optimal codons (Fop), the effective number of codons (NC), and the GC content of synonymous third codon positions (GC3s)¹¹¹. To discover mutation hotspot sites, the nucleotide diversity in the *Lilium* chloroplast genomes was quantified employing sliding window analysis via DnaSP 6 software¹¹². The window length and step size were fixed at 600 bp and 200 bp, respectively.

Selection pressure on *Lilium* cp genomes. We extracted the CDS sequences of the protein-coding genes from all 48 species, and flushed out those with lacking data in at least one species, which resulted in 78 CDS matrices. MAFFT v7¹¹³ was used to generate CDS alignments. DnaSP 6¹¹² was used to compute the rates of nonsynonymous (Ka) and synonymous (Ks) nucleotide substitution. The selective pressure was measured using the Ka/Ks ratio, with Ka/Ks < 1, Ka/Ks = 1, and Ka/Ks > 1, indicating purifying, neutral, and positive selection, respectively⁶⁷.

Phylogenetic analysis. The MAFFT v7 program was employed to align the complete cp genome sequence of 47 *Lilium* species¹¹³. Two complete sequences of *Fritillaria hupehensis* and *Fritillaria cirrhosa* were applied as outgroups. Utilizing the effective nucleotide substitution model (GTR + G), the maximum likelihood was carried out with RAxML v8.2.11¹¹⁴ with 1000 bootstrap repetitions to construct the phylogenetic tree. Finally, the iTOL tool was employed for visualizing the coming about phylogenetic tree¹¹⁵. Furthermore, phylogenetic analyses were also carried out for the single-copy protein coding genes (CDSs) shared by all 47 species.

Received: 1 December 2021; Accepted: 24 May 2022

Published online: 07 June 2022

References

- Chase, M. W. *et al.* An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **181**, 1–20 (2016).
- Rudall, P. J. *et al.* Consider the lilies: Systematics of Liliales. *Monocots Syst. Evol.* **1**, 347–359 (2000).
- Kim, J. S., Hong, J.-K., Chase, M. W., Fay, M. F. & Kim, J.-H. Familial relationships of the monocot order Liliales based on a molecular phylogenetic analysis using four plastid loci: matK, rbcL, atpB and atpF-H. *Bot. J. Linn. Soc.* **172**, 5–21 (2013).
- Givnish, T. J. *et al.* Phylogenomics and historical biogeography of the monocot order Liliales: Out of Australia and through Antarctica. *Cladistics* **32**, 581–605 (2016).
- Group, A. P. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* **161**, 105–121 (2009).
- Petersen, G., Seberg, O. & Davis, J. I. Phylogeny of the Liliales (Monocotyledons) with special emphasis on data partition congruence and RNA editing. *Cladistics* **29**, 274–295 (2013).
- Peruzzi, L. A new infrafamilial taxonomic setting for Liliaceae, with a key to genera and tribes. *Plant Biosyst. Int. J. Deal. Asp. Plant Biol.* **150**, 1341–1347 (2016).
- Kim, J. S. & Kim, J.-H. Updated molecular phylogenetic analysis, dating and biogeographical history of the lily family (Liliales). *Bot. J. Linn. Soc.* **187**, 579–593 (2018).
- Li, J. *et al.* Phylogeny, age, and evolution of tribe Liliae (Liliales) based on whole plastid genomes. *Front. Plant Sci.* **12**, 699226 (2021).
- MacRae, E. A. *Lilies: A Guide for Growers and Collectors* (Timber Press, 1998).
- Comber, H. F. A new classification of the genus *Lilium*. *Lily Year B. RHS* 86–105 (1949).

12. Kim, H. T., Lim, K.-B. & Kim, J. S. New insights on *Lilium* phylogeny based on a comparative phylogenomic study using complete plastome sequences. *Plants* **8**, 547 (2019).
13. Nishikawa, T., Okazaki, K., Uchino, T., Arakawa, K. & Nagamine, T. A molecular phylogeny of *Lilium* in the internal transcribed spacer region of nuclear ribosomal DNA. *J. Mol. Evol.* **49**, 238–249 (1999).
14. İkinçi, N., Oberprieler, C. & Güner, A. On the origin of European lilies: Phylogenetic analysis of *Lilium* section *Liriotypus* (Liliaceae) using sequences of the nuclear ribosomal transcribed spacers. *Willdenowia* **36**, 647–656 (2006).
15. Muratović, E., Hidalgo, O., Garnatje, T. & Siljak-Yakovlev, S. Molecular phylogeny and genome size in European lilies (Genus *Lilium*, Liliaceae). *Adv. Sci. Lett.* **3**, 180–189 (2010).
16. Gong, X. *et al.* Frequent gene flow blurred taxonomic boundaries of sections in *Lilium* L. (Liliaceae). *PLoS ONE* **12**, e0183209 (2017).
17. Kim, J.-H. *et al.* Chloroplast genomes of *Lilium lancifolium*, *L. amabile*, *L. callosum*, and *L. philadelphicum*: Molecular characterization and their use in phylogenetic analysis in the genus *Lilium* and other allied genera in the order Liliales. *PLoS ONE* **12**, e0186788 (2017).
18. Du, Y. *et al.* Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses. *Sci. Rep.* **7**, 1–10 (2017).
19. Van Tuyl, J. M. *et al.* *Lilium*. In *Ornamental Crops* 481–512 (Springer, 2018).
20. Wendelbo, P. Tulips and Irises of Iran and their relatives. (1977).
21. Bakhshaie, M., Babalar, M., Mirmasoumi, M. & Khalighi, A. Somatic embryogenesis and plant regeneration of *Lilium ledebourii* (Baker) Boiss., an endangered species. *Plant Cell Tissue Organ Cult.* **102**, 229–235 (2010).
22. Azadi, P. & Khosh-Khui, M. Micropropagation of *Lilium ledebourii* (Baker) Boiss as affected by plant growth regulator, sucrose concentration, harvesting season and cold treatments. *Electron. J. Biotechnol.* **10**, 582–591 (2007).
23. Rechinger, K. H. Compositae VII: *Calendula*. In *Flora Iranica* no. 164: 99–105, *Graz Akad. Druck-und Verlagsanstalt* (ed Rechinger, K. H.) (1989).
24. Kavian, B. Micromorphological, morphological and anatomical investigation of the *Lilium ledebourii* (Baker) Boiss. (Liliaceae) indigenous to Iran. *Aust. J. Cr. Sci.* **1**, 6–10 (2008).
25. Bakhshaie, M., Babalar, M., Mirmasoumi, M. & Khalighi, A. Effects of light, sucrose, and cytokinins on somatic embryogenesis in *Lilium ledebourii* (Baker) Boiss. via transverse thin cell-layer cultures of bulblet microscales. *J. Hortic. Sci. Biotechnol.* **85**, 491–496 (2010).
26. Sawarkar, A. D. *et al.* Traditional system versus DNA barcoding in identification of bamboo species: A systematic review. *Mol. Biotechnol.* **63**, 651–675 (2021).
27. Jansen, R. K. *et al.* Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci.* **104**, 19369–19374 (2007).
28. Zhao, K. *et al.* Comparative analyses of chloroplast genomes from 14 *Zanthoxylum* species: Identification of variable DNA markers and phylogenetic relationships within the genus. *Front. Plant Sci.* **11**, 2226 (2021).
29. Bobik, K. & Burch-Smith, T. M. Chloroplast signaling within, between and beyond cells. *Front. Plant Sci.* **6**, 781 (2015).
30. Neuhaus, H. E. & Emes, M. J. Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Biol.* **51**, 111–140 (2000).
31. Pfannschmidt, T. & Yang, C. The hidden function of photosynthesis: A sensing system for environmental conditions that regulates plant acclimation responses. *Protoplasma* **249**, 125–136 (2012).
32. Zhang, T. *et al.* The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: Insights into the evolution of plant organellar genomes. *PLoS ONE* **7**, e30531 (2012).
33. Wolfe, K. H., Li, W.-H. & Sharp, P. M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci.* **84**, 9054–9058 (1987).
34. Twyford, A. D. & Ness, R. W. Strategies for complete plastid genome sequencing. *Mol. Ecol. Resour.* **17**, 858–868 (2017).
35. Olmstead, R. G. & Palmer, J. D. Chloroplast DNA systematics: A review of methods and data analysis. *Am. J. Bot.* **81**, 1205–1224 (1994).
36. Khan, A. L., Asaf, S., Al-Rawahi, A. & Al-Harrasi, A. Decoding first complete chloroplast genome of toothbrush tree (*Salvadora persica* L.): Insight into genome evolution, sequence divergence and phylogenetic relationship within Brassicales. *BMC Genomics* **22**, 1–16 (2021).
37. Nguyen, H. Q. *et al.* Complete chloroplast genome of novel *Adrinandra megaphylla* Hu species: Molecular structure, comparative and phylogenetic analysis. *Sci. Rep.* **11**, 1–11 (2021).
38. Zhu, B. *et al.* Complete chloroplast genome features and phylogenetic analysis of *Eruca sativa* (Brassicaceae). *PLoS ONE* **16**, e0248556 (2021).
39. Dong, W., Liu, J., Yu, J., Wang, L. & Zhou, S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* **7**, e35071 (2012).
40. Li, W., Zhang, C., Guo, X., Liu, Q. & Wang, K. Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis. *PLoS ONE* **14**, e0216645 (2019).
41. Park, I. *et al.* The complete chloroplast genomes of six *Ipomoea* species and indel marker development for the discrimination of authentic *Pharbitidis* Semen (Seeds of *I. nil* or *I. purpurea*). *Front. Plant Sci.* **9**, 965 (2018).
42. Zhang, W., Wang, H., Dong, J., Zhang, T. & Xiao, H. Comparative chloroplast genomes and phylogenetic analysis of *Aquilegia*. *Appl. Plant Sci.* **9**, e11412 (2021).
43. Wen, F. *et al.* The complete chloroplast genome of *Stauntonia chinensis* and compared analysis revealed adaptive evolution of subfamily Lardizabaloideae species in China. *BMC Genomics* **22**, 1–18 (2021).
44. Xue, J., Wang, S. & Zhou, S. Polymorphic chloroplast microsatellite loci in *Nelumbo* (Nelumbonaceae). *Am. J. Bot.* **99**, e240–e244 (2012).
45. Ogihara, Y., Terachi, T. & Sasakuma, T. Intramolecular recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species. *Proc. Natl. Acad. Sci.* **85**, 8573–8577 (1988).
46. Bulmer, M. Are codon usage patterns in unicellular organisms determined by selection-mutation balance? *J. Evol. Biol.* **1**, 15–26 (1988).
47. Rouwendal, G. J. A., Mendes, O., Wolbert, E. J. H. & De Boer, A. D. Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage. *Plant Mol. Biol.* **33**, 989–999 (1997).
48. Liu, H.-Y. *et al.* The chloroplast genome of *Lilium henrici*: Genome structure and comparative analysis. *Molecules* **23**, 1276 (2018).
49. Bock, R. & Knoop, V. *Genomics of Chloroplasts and Mitochondria* 35 (Springer Science & Business Media, 2012).
50. Zhu, A., Guo, W., Gupta, S., Fan, W. & Mower, J. P. Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. *New Phytol.* **209**, 1747–1756 (2016).
51. Dong, W.-L. *et al.* Molecular evolution of chloroplast genomes of orchid species: Insights into phylogenetic relationship and adaptive evolution. *Int. J. Mol. Sci.* **19**, 716 (2018).
52. Li, P. *et al.* Comparative genomics and phylogenomics of East Asian tulips (*Amana*, Liliaceae). *Front. Plant Sci.* **8**, 451 (2017).
53. Chen, Q., Wu, X. & Zhang, D. Phylogenetic analysis of *Fritillaria cirrhosa* D. Don and its closely related species based on complete chloroplast genomes. *PeerJ* **7**, e7480 (2019).
54. Lu, R.-S., Li, P. & Qiu, Y.-X. The complete chloroplast genomes of three *Cardiocrinum* (Liliaceae) species: Comparative genomic and phylogenetic analyses. *Front. Plant Sci.* **7**, 2054 (2017).

55. Zhou, L. *et al.* Developing single nucleotide polymorphism markers for the identification of pineapple (*Ananas comosus*) germplasm. *Hortic. Res.* **2**, 1–12 (2015).
56. Wang, X. *et al.* The USDA cucumber (*Cucumis sativus* L.) collection: Genetic diversity, population structure, genome-wide association studies, and core collection development. *Hortic. Res.* **5**, 1–13 (2018).
57. Kong, B.L.-H. *et al.* Comparative analysis and phylogenetic investigation of Hong Kong Ilex chloroplast genomes. *Sci. Rep.* **11**, 1–13 (2021).
58. Asano, T., Tsudzuki, T., Takahashi, S., Shimada, H. & Kadowaki, K. Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: A comparative analysis of four monocot chloroplast genomes. *DNA Res.* **11**, 93–99 (2004).
59. Liu, S., Wang, Z., Su, Y. & Wang, T. Comparative genomic analysis of Polypodiaceae chloroplasts reveals fine structural features and dynamic insertion sequences. *BMC Plant Biol.* **21**, 1–15 (2021).
60. Huang, J. *et al.* Comparative chloroplast genomics of Fritillaria (Liliaceae), inferences for phylogenetic relationships between Fritillaria and Lilium and plastome evolution. *Plants* **9**, 133 (2020).
61. Maréchal, A. & Brisson, N. Recombination and the maintenance of plant organelle genome stability. *New Phytol.* **186**, 299–317 (2010).
62. Yuan, C. *et al.* Uncovering dynamic evolution in the plastid genome of seven Ligusticum species provides insights into species discrimination and phylogenetic implications. *Sci. Rep.* **11**, 1–13 (2021).
63. Khakhlova, O. & Bock, R. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* **46**, 85–94 (2006).
64. Huo, Y. *et al.* Complete chloroplast genome sequences of four Allium species: Comparative and phylogenetic analyses. *Sci. Rep.* **9**, 1–14 (2019).
65. Li, J. *et al.* Comparative analysis of the complete plastid genome of five Bupleurum species and new insights into DNA barcoding and phylogenetic relationship. *Plants* **9**, 543 (2020).
66. Choi, K. S., Chung, M. G. & Park, S. The complete chloroplast genome sequences of three Veroniceae species (Plantaginaceae): Comparative analysis and highly divergent regions. *Front. Plant Sci.* **7**, 355 (2016).
67. Kimura, M. The neutral theory of molecular evolution and the world view of the neutralists. *Genome* **31**, 24–31 (1989).
68. Fay, J. C. & Wu, C.-I. Sequence divergence, functional constraint, and selection in protein evolution. *Annu. Rev. Genomics Hum. Genet.* **4**, 213–235 (2003).
69. Sablok, G., Mudunuri, S. B., Edwards, D. & Ralph, P. J. Chloroplast genomics: Expanding resources for an evolutionary conserved miniature molecule with enigmatic applications. *Curr. Plant Biol.* **7**, 34–38 (2016).
70. Rogalski, M., Ruf, S. & Bock, R. Tobacco plastid ribosomal protein S18 is essential for cell survival. *Nucleic Acids Res.* **34**, 4537–4545 (2006).
71. Liu, M.-L. *et al.* Evolutionary analysis of plastid genomes of seven *Lonicera* L. species: Implications for sequence divergence and phylogenetic relationships. *Int. J. Mol. Sci.* **19**, 4039 (2018).
72. Chen, S. L. & Xiao, P. G. Molecular evolution and positive Darwinian selection of the chloroplast maturase matK. *J. Plant Res.* **123**, 241–247 (2010).
73. Peltier, G., Aro, E.-M. & Shikanai, T. NDH-1 and NDH-2 plastoquinone reductases in oxygenic photosynthesis. *Annu. Rev. Plant Biol.* **67**, 55–80 (2016).
74. Yamori, W. & Shikanai, T. Physiological functions of cyclic electron transport around photosystem I in sustaining photosynthesis and plant growth. *Annu. Rev. Plant Biol.* **67**, 81–106 (2016).
75. Drescher, A., Ruf, S., Calsa, T. Jr., Carrer, H. & Bock, R. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant J.* **22**, 97–104 (2000).
76. Kuroda, H. & Maliga, P. The plastid clpP1 protease gene is essential for plant development. *Nature* **425**, 86–89 (2003).
77. Shi, H. *et al.* Complete chloroplast genomes of two *Siraitia* Merrill species: Comparative analysis, positive selection and novel molecular marker development. *PLoS ONE* **14**, e0226865 (2019).
78. Kapralov, M. V. & Filatov, D. A. Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evol. Biol.* **7**, 1–10 (2007).
79. Hudson, G. S. & Mason, J. G. The chloroplast genes encoding subunits of the H⁺-ATP synthase. In *Mol. Biol. Photosynth.* 565–582 (1988).
80. Piot, A., Hackel, J., Christin, P.-A. & Besnard, G. One-third of the plastid genes evolved under positive selection in PACMAD grasses. *Planta* **247**, 255–266 (2018).
81. Jalal, A. *et al.* A small multifunctional pentatricopeptide repeat protein in the chloroplast of *Chlamydomonas reinhardtii*. *Mol. Plant* **8**, 412–426 (2015).
82. Zhou, T. *et al.* Complete chloroplast genome sequence determination of Rheum species and comparative chloroplast genomics for the members of Rumiceae. *Plant Cell Rep.* **39**, 811–824 (2020).
83. Raman, G. & Park, S. The complete chloroplast genome sequence of the *Speirantha gardenii*: Comparative and adaptive evolutionary analysis. *Agronomy* **10**, 1405 (2020).
84. Gao, L.-Z. *et al.* Evolution of *Oryza* chloroplast genomes promoted adaptation to diverse ecological habitats. *Commun. Biol.* **2**, 1–13 (2019).
85. Hurst, L. D. The Ka/Ks ratio: Diagnosing the form of sequence evolution. *Trends Genet. TIG* **18**, 486 (2002).
86. Zhou, T. *et al.* Comparative chloroplast genome analyses of species in Gentiana section Cruciata (Gentianaceae) and the development of authentication markers. *Int. J. Mol. Sci.* **19**, 1962 (2018).
87. Tiffin, P. & Hahn, M. W. Coding sequence divergence between two closely related plant species: *Arabidopsis thaliana* and *Brassica rapa* ssp. *pekinensis*. *J. Mol. Evol.* **54**, 746–753 (2002).
88. Zhang, J. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J. Mol. Evol.* **50**, 56–68 (2000).
89. Raven, J. A., Beardall, J., Larkum, A. W. D. & Sánchez-Baracaldo, P. Interactions of photosynthesis with genome size and function. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20120264 (2013).
90. Kumar, R. A., Oldenburg, D. J. & Bendich, A. J. Changes in DNA damage, molecular integrity, and copy number for plastid DNA and mitochondrial DNA during maize development. *J. Exp. Bot.* **65**, 6425–6439 (2014).
91. Rohde, K. Latitudinal gradients in species diversity: the search for the primary cause. *Oikos* 514–527 (1992).
92. Du, Y. *et al.* Molecular phylogeny and genetic variation in the genus *Lilium* native to China based on the internal transcribed spacer sequences of nuclear ribosomal DNA. *J. Plant Res.* **127**, 249–263 (2014).
93. Nishikawa, T., Okazaki, K., Arakawa, K. & Nagamine, T. Phylogenetic analysis of section *Sinomartagon* in genus *Lilium* using sequences of the internal transcribed spacer region in nuclear ribosomal DNA. *Breed. Sci.* **51**, 39–46 (2001).
94. Gao, Y., Zhou, S. & He, X. Karyotype studies in thirty-two species of *Lilium* (Liliaceae) from China 1. *Nord. J. Bot.* **29**, 746–761 (2011).
95. Gao, Y.-D., Harris, A. J. & He, X.-J. Morphological and ecological divergence of *Lilium* and *Nomocharis* within the Hengduan Mountains and Qinghai-Tibetan Plateau may result from habitat specialization and hybridization. *BMC Evol. Biol.* **15**, 1–21 (2015).
96. Gao, Y.-D., Harris, A. J., Zhou, S.-D. & He, X.-J. Evolutionary events in *Lilium* (including *Nomocharis*, Liliaceae) are temporally correlated with orogenies of the Q-T plateau and the Hengduan Mountains. *Mol. Phylogenet. Evol.* **68**, 443–460 (2013).

97. Gao, Y.-D., Zhou, S.-D. & He, X.-J. *Lilium yapingense* (Liliaceae), a new species from Yunnan, China, and its systematic significance relative to *Nomocharis*. In *Annales Botanici Fennici* Vol. 50, 187–194 (BioOne, 2013).
98. Zhou, Y.-P. *et al.* Fluorescence in situ hybridization of 35S rDNA sites and karyotype of wild *Lilium* (Liliaceae) species from China: Taxonomic and phylogenetic implications. *Genet. Resour. Crop Evol.* **67**, 1601–1617 (2020).
99. Ghanbari, S., Fakheri, B. A., Naghavi, M. R. & Mahdinezhad, N. Evaluating phylogenetic relationships in the *Lilium* family using the ITS marker. *J. Plant Biotechnol.* **45**, 236–241 (2018).
100. Tokgun, O., Akca, H., Mammadov, R., Aykurt, C. & Deniz, G. *Convulvulus galaticus*, *Crocus antalyensis*, and *Lilium candidum* extracts show their antitumor activity through induction of p53-mediated apoptosis on human breast cancer cell line MCF-7 cells. *J. Med. Food* **15**, 1000–1005 (2012).
101. Chaisson, M. J. & Tesler, G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory. *BMC Bioinformatics* **13**, 1–18 (2012).
102. Miyamoto, M. *et al.* Performance comparison of second- and third-generation sequencers using a bacterial genome with two chromosomes. *BMC Genomics* **15**, 1–9 (2014).
103. Soorni, A., Haak, D., Zaitlin, D. & Bombarely, A. Organelle_PBA, a pipeline for assembling chloroplast and mitochondrial genomes from PacBio DNA sequencing data. *BMC Genomics* **18**, 1–8 (2017).
104. Tillich, M. *et al.* GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**, W6–W11 (2017).
105. Chan, P. P. & Lowe, T. M. tRNAscan-SE: searching for tRNA genes in genomic sequences. In *Gene Prediction* 1–14 (Springer, 2019).
106. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **47**, W59–W64 (2019).
107. Beier, S., Thiel, T., Münch, T., Scholz, U. & Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **33**, 2583–2585 (2017).
108. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
109. Kurtz, S. The Vmatch large scale sequence analysis software—A manual. *Cent. Bioinform.* **170**, 391–392 (2010).
110. Xia, X. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. *Mol. Biol. Evol.* **35**, 1550–1552 (2018).
111. Xie, D.-F. *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.* **10**, 460 (2019).
112. Rozas, J. *et al.* DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
113. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
114. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
115. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

Author contributions

The study was conceived and designed by M.S., R.N., and M.K. The samples were collected by M.S., R.N., and M.K. M.S. performed all data analysis and wrote the manuscript. The research was supervised by R.N., M.F.M. and V.S. R.N., S.A.S., M.F.M. and V.S. revised the manuscript. All authors contributed to the article. The final manuscript was read and approved by the authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-13449-x>.

Correspondence and requests for materials should be addressed to M.S.-A. or R.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022