RESEARCH



The evolution of ephemeral flora in Xinjiang, China: insights from plastid phylogenomic analyses of Brassicaceae

Tian-Wen Xiao¹, Feng Song¹, Duc Quy Vu¹, Ying Feng² and Xue-Jun Ge^{1*}

Abstract

Background The ephemeral flora of northern Xinjiang, China, plays an important role in the desert ecosystems. However, the evolutionary history of this flora remains unclear. To gain new insights into its origin and evolutionary dynamics, we comprehensively sampled ephemeral plants of Brassicaceae, one of the essential plant groups of the ephemeral flora.

Results We reconstructed a phylogenetic tree using plastid genomes and estimated their divergence times. Our results indicate that ephemeral species began to colonize the arid areas in north Xinjiang during the Early Miocene and there was a greater dispersal of ephemeral species from the surrounding areas into the ephemeral community of north Xinjiang during the Middle and Late Miocene, in contrast to the Early Miocene or Pliocene periods.

Conclusions Our findings, together with previous studies, suggest that the ephemeral flora originated in the Early Miocene, and species assembly became rapid from the Middle Miocene onwards, possibly attributable to global climate changes and regional geological events.

Keywords Ephemeral flora, Brassicaceae, Species assembly, Divergence time

Background

Ephemerals are plants that inhabit in arid regions, relying on rainfall and snowmelt water during spring and completing their life cycles within approximately two months before the onset of summer. They are also termed spring annuals, short-trophophase plants, short-living plants, or early-spring ephemeral plants [1, 2], and are typically found in North America, Western and Central Asia, the Mediterranean region, and Northern and Southern

*Correspondence:

Xue-Jun Ge

xjge@scbg.ac.cn

¹ Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

² State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China

Africa, with Central Asia being the distribution center [2, 3]. In China, ephemeral flora is mainly distributed in northern Xinjiang, particularly the Junggar Basin and its adjacent regions (Fig. 1), and is an important component of the Central Asian flora. In this region, there are 207 ephemeral species, forming 97 genera and 27 families, and covering 6.5% of the total species found in the Xinjiang floras [2]. Among the 27 families, Liliaceae harbors the largest number of ephemeral plants (37 species), followed by Brassicaceae (33 species), Boraginaceae (17 species), Fabaceae (15 species), Asteraceae (14 species), Apiaceae (13 species), and Poaceae (11 species) [2, 3].

Ephemeral flora plays an important role in desert ecosystems. For example, it makes a major contribution to land fixation in the Gurbantunggut Desert [4]; it improves the soil quality in the desert-oasis ecotone [5]; and it is an important feed source for grass-feeding livestock in early spring [2]. Despite its importance,



© The Author(s) 2024. **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.gr/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.gr/licenses/by/4.0/. The Creative Commons Public Domain Dedicated in a credit line to the data.



Fig. 1 Geographic location of the studying area

ephemeral flora faces the threat of overexploitation and climate change. The assembly of a flora is a complex process spanning over large time scales, and is influenced by environmental conditions, physiological properties, and evolutionary histories of plants [6, 7]. Therefore, understanding the historical dynamics of species assembly of ephemeral flora can provide insights into its future biodiversity in a rapidly changing world. However, the origin and evolutionary history of ephemeral flora in northern Xinjiang remains unclear.

Mao and Zhang [3] proposed that the ephemeral flora in Xinjiang only occurred after the disappearance of the Paratethys Sea and originated from xerothermic vegetation around the Pliocene-Pleistocene transition. However, this viewpoint lacks supporting evidence from paleontology or well-dated phylogeny. Li et al. [8] estimated the divergence times of Brassicaceae ephemeral species using trnL-trnF and ITS and inferred that ephemeral flora originated during the Middle and Late Miocene (14–6 Mya). Nevertheless, the estimation of the divergence time may be biased due to the lack of sufficient parsimony-informative sites within several molecular markers [9, 10]. Additionally, they did not consider the origination times of ephemeral plants from the other families [8]. Therefore, the origin and evolution of ephemeral flora in northern Xinjiang require further investigation.

Brassicaceae harbors the second largest number of ephemeral species in the ephemeral flora of northern Xinjiang (Fig. 2), which are dominant or companion species in plant communities. Furthermore, ephemeral



Fig. 2 Ephemeral species of Brassicaceae. **a** Habitat; **b** Ephemeral plant community; **c** *Chorispora tenella*; **d** *Goldbachia sabulosa*; **e** *Isatis gymnocarpa*; **f** *Isatis minima*; **g** *Isatis multicaulis*; **h** *Lachnoloma lehmannii*. **a–c** Photographed by Ying Feng, and **d–h** by Xin-Xin Zhou

plants of Brassicaceae belong to 22 genera, which are larger than the other families [2]. Previous studies have reported hundreds of plastomes of Brassicaceae, including ephemeral and non-ephemeral species, and have shown well-resolved phylogenies [11–14], which provide a solid foundation for further investigating the origin and diversification of ephemeral plants. Considering all these factors, Brassicaceae represents an ideal group of plants for studying the evolutionary dynamics of ephemeral flora.

In this study, the species names of ephemeral plants in Brassicaceae were collected from Ephemeral Plants in *Xinjiang, China* [2], and standardized using the Plants of the World Online (POWO). As a result, a list of names belonging to 32 ephemeral species from 21 genera was obtained (Table S1). Sixteen ephemeral species of Brassicaceae were sampled and their plastomes were sequenced; the sequencing data from these 16 ephemeral species were combined with plastomes from another eight ephemeral species from GenBank. Thus, finally a total of 24 (75%) ephemeral species were included in this study (Table S1). Based on these data, this study aimed to characterize the structural variation of plastomes, infer the positions of ephemeral plants in the Brassicaceae phylogenetic tree, and estimate their divergence times, in the hope of providing insights into the evolutionary dynamics of ephemeral flora.

Materials and methods

Taxon sampling and DNA extraction

Forty-nine samples representing 40 species and 20 genera of Brassicaceae were collected from Xinjiang and Gansu, China, and identified by Ying Feng and Yan Li. Among these species, 16 were ephemeral (Table S2). Voucher specimens were deposited in the herbarium of the South China Botanical Garden of the Chinese Academy of Sciences (IBSC). Silica gel-dried leaf tissues were used for DNA extraction using the cetyltrimethylammonium bromide (CTAB) method [15]. Genomic DNA concentration was determined using the Qubit 3.0 Fluorometer dsDNA HS Assay Kit (Invitrogen, Carlsbad, CA, USA). No specific permissions or licenses were required for the collections and experiments.

Plastome sequencing, assembly, and annotation

Library preparation and genome skimming sequencing were performed at the Beijing Genomics Institute (BGI, Shenzhen) following the method as described by Liu et al. [16]. For each sample, 1 μ g of genomic DNA was randomly fragmented into small pieces using Covaris (Covaris, USA) and fragments of 200–400 bp were selected for PCR amplification. The amplified sequences were purified using the Agencourt AMPure XP-Medium kit (Avantor, USA). The final library was qualified using the Agilent Technologies 2100 bioanalyzer (Agilent DNA 1000 Reagents), and sequenced on the BGISEQ-500 platform (paired-end 150bp). Approximately 2–3 Gb raw data were obtained for each sample. Quality and length filtering, adapter trimming, and quality check were performed using fastp v0.23.2 with default parameters [17].

Plastomes were assembled using GetOrganelle v1.7.5.3 [18]. To ensure that the plastomes were correctly assembled, clean reads were mapped on plastomes using Burrows-Wheeler Aligner v0.7.17-r1188 [19], converted to bam file using SAMtools v1.9 [20], and manually inspected in Geneious v9.1.3 [21]. Plastomes were annotated using the online program GeSeq [22]. The annotations were then compared to plastomes of the same genus downloaded from GenBank and corrected when necessary. The precise locations of the start and stop codons were checked and adjusted in Geneious v9.1.3. Linear plastome maps were generated using OGDRAW v1.3.1 [23]. Raw sequence reads and assembled plastomes were submitted to the Sequence Read Archive (SRA) of NCBI and GenBank (Table S2).

Plastome feature analyses

The expansion and contraction of the large single copy (LSC), small single copy (SSC) and inverted repeat (IR) regions of newly sequenced plastomes were visualized using the IRscope v0.1 [24]. To detect dispersed repeats (including forward, reverse, complement, and palindromic repeats) in each plastome, the online program REPuter was used with default settings [25]. Simple sequence repeats (SSRs) were determined using the MIcroSAtellite identification tool (MISA v2.1) [26] with all parameters set following Xiao and Ge [9]. Tandem repeats were detected using the online program Tandem Repeats Finder v4.09 [27] with default parameters. To explore the contribution of repeat number and maximum length to plastome length and GC content variation, the generalized linear model was employed to calculate the coefficients and p values in R v4.0.4 [28].

Before sequence alignment, the direction of the reversed segments was manually adjusted. The 49 newly sequenced plastomes were aligned using MAFFT v7.508 [29] with default parameters. To identify hypervariable regions, nucleotide diversity (Pi) values were calculated using DnaSP v5.10.01 [30]. The window length and step size were set as 800 and 200, respectively. The Pi value of each site was plotted using ggplot2 [31] in R v4.0.4.

Phylogenetic analyses

One hundred and sixty-three plastomes representing all major clades of Brassicaceae [11] and one plastome of *Cleome chrysantha* were downloaded from GenBank for

maximum likelihood (ML) tree inference (Table S3). All loci of the 164 downloaded and 49 newly sequenced plastomes were extracted using a python script (get_annotated_regions_from_gb.py [32]). The protein-coding genes (PCGs) and non- protein-coding genes (including tRNAs, rRNAs, introns, pseudogenes, and intergenic spacers) were separately aligned using MAFFT under the localpair mode and with 1000 iterative refinements. To remove poorly aligned regions and improve the quality of subsequent analyses, alignments were trimmed using trimAl v1.4 [33] with the "-automated1" flag. The aligned loci were concatenated using AMAS v1.0 [34], generating three sequence matrices, i.e., the concatenated PCGs (PCGs-con), the concatenated non-PCGs (NPCGs-con), and complete plastomes with one IR removed (CP-con). The alignment lengths, number of variable sites, number of parsimony-informative sites, and GC content of the three matrices were summarized using AMAS v1.0 [34].

For the three matrices, ML tree construction was performed using RAxML v8.2.11 [35] with the GTR-GAMMA model and 1000 rapid bootstrap replicates. Because the partitioned strategy of sequence data can improve the accuracy of tree inference [36], data partitioning was applied in this study. Specifically, each locus was treated as an independent block, and the best partition scheme was determined by ModelFinder [37] implemented in IQ-Tree v1.6.8 [38]. ML analysis in IQ-Tree was performed with 1000 ultrafast bootstraps (UFBS) [39] and 1000 Shimodaira-Hasegawa-like approximate likelihood ratio tests (SH-aLRTs) [40]. To reduce the computational cost, the partitioned strategy was applied only to the PCGs-con. All trees were visualized using FigTree [41].

Divergence time estimation

To trace the evolutionary history of ephemeral plants in Brassicaceae, molecular dating was performed using a penalized-likelihood method implemented in treePL v1.0 [42]. Before the analysis, 23 plastomes representing Vitales, Malpighiales, Fabales, Cucurbitales, Fagales, Rosales, Myrtales, Sapindales, Mavales, and Brassicales were downloaded from GenBank as outgroups of Brassicaceae (Table S3). Loci extraction, aligning, trimming, and concatenation were performed as described in the above section "Phylogenetic analyses". Two datasets, i.e., concatenated PCGs (PCGs-con-div) and complete plastomes with one IR removed (CP-con-div), were generated for divergence time estimation. The two datasets were used for ML analyses in RAxML v8.2.11 with the GTRGAMMA model 1000 bootstrap replicates.

Four fossil calibrations were chosen for divergence time estimation following the methods described by Hohmann et al. [13], Huang et al. [43], and Walden et al. [14]. The minimum age for the splits of *Citrus/Mangifera*, *Oenothera/Eucalyptus*, *Prunus/Malus*, and *Castanea/Cucumis* was set to 65, 88.2, 48.4, and 84 Mya, respectively. The maximum age of the four calibrations was set to 125 Mya. The root age was constrained to a minimum age of 92 Mya and a maximum age of 125 Mya according to the estimation of Magallón et al. [44]. The fossil *Thlaspi primaevum* from Brassicaceae is still under debate; therefore, it was not included in the present study [45].

The 1000 bootstrap trees of PCGs-con-div and CPcon-div were used as inputs in treePL. To determine the appropriate level of rate heterogeneity in the phylograms, random sampled cross-validation was conducted to obtain the optimal smoothing value for each tree. The parameters cvstart and cvstop were set to 100,000 and 0.001, respectively, while the other parameters were set to default. The output trees were then used to generate the time tree by TreeAnnotator implemented in BEAST v2.6.0 [46].

In addition, divergence times were estimated using the Bayesian method MCMCtree implemented in PAML v4.9j [47], which allows soft bounds for fossil calibrations and uses the Compound Dirichlet prior for nucleotide substitution rates. The best-scoring ML tree inferred from PCGs-con-div was used as input, and fossil calibrations were set following treePL analysis. The gradient and Hessian were calculated using the MCMCtree and BASEML programs in PAML, and the output was used as input in the next step. Thereafter, MCMC sampling was performed to obtain the posterior distribution using the approximate likelihood method with the following parameters: model was set as HKY85, rgene gamma as 1 2 1, and sigma2_gamma as 1 10 1. After a burn-in of the first 20,000 generations, the MCMC run was sampled every 100 generations until 10,000 samples were collected. Two MCMC runs were performed with different random seeds, and convergence was checked in Tracer v1.7.1 [48].

Substitution rate variation

To detect the substitution rate variation between ephemeral and non-ephemeral plants, the substitution rate of each species was calculated as r = d/2T, while r was substitution rate, d was substitutions per site, and T was the divergence time. The substitutions per site (tip-to-root distance) for each species was extracted from ML tree using PhyKit v1.11.15 [49] with *Cleome chrysantha* set as the root.

Phylogenetic signal

Ancestral state reconstruction (ASR) is commonly used to infer the evolutionary history of a trait; however, it is recommended that ASR should be performed on trees with strong phylogenetic signals to obtain accurate reconstructions [50]. Therefore, the Blomberg K and Pagel's λ were calculated using the phylosig function in the R package phytools [51]. Before the analysis, ephemeral and non-ephemeral plants were coded as 1 and 2, respectively.

Results

Features of newly sequenced plastomes

In this study, 49 complete plastomes were generated, which all displayed a typical quadripartite structure (i.e., LSC, IRb, SSC, and IRa). The complete plastome lengths ranged from 150,682 bp (Alyssum simplex HM2130) to 162,956 bp (Chorispora sibirica HM489), LSC from 80,743 bp (Alyssum simplex HM2130) to 86,590 bp (Chorispora sibirica HM489), IR from 26,062 bp (Eutrema nepalense HM0150) to 32,908 bp (Chorispora sibirica HM489), and SSC from 10,523 bp (Chorispora sibirica HM2158) to 18,172 bp (Draba rockii HM0131) (Table 1). Gene content of tRNA and rRNA was conserved, each containing 30 unique tRNAs and four unique rRNAs (Table 1). However, rps16, ycf15 and accD were pseudolized in 15, two, and two plastomes, respectively (Table 1). The overall GC content was 35.9%-36.6%. Notably, ycf2, ycf15, trnL-UUG, and their flanking intergenic spacers were only inverted in Chorispora sibirica HM489 and HM2158 (Fig. 3).

In the plastomes of Chorispora sibirica HM489 and HM2158, the complete ycf1, rps15, and ndhH doubled in the IR regions, which contributed to the extreme IR expansion toward the SSC region (Figs. 3 and S1). To ensure that the expansion was not caused by sequencing errors or misassembly, clean reads of the two samples were mapped to plastomes and inspected in Geneious. The mapping results showed that IR expansion occurred in the two plastomes of Chorispora sibirica (i.e., HM489 and HM2158) (Fig. S2), but not in the other 47 plastomes. In addition, the IR regions of Chorispora sibirica HM489 and HM2158 shrank slightly at the LSC/Irb boundary that the complete *rpl2* gene was only partially present in the IR regions of Chorispora sibirica HM489 and HM2158 (Figs. 3 and S1), but fully present in the IR regions of the other 47 plastomes.

The repeats and hypervariable regions

The number of palindromic repeats was generally higher than that of forward repeats, followed by reverse and complement repeats (Table S4). The maximum length of dispersed repeats of *Chorispora sibirica* HM489 and HM2158 were 281 bp and 214 bp, respectively, which were larger than that of the other plastomes (\leq 96 bp) except *Sisymbrium loeselii* HM2047 (185 bp). For the

SSRs analysis, mono-, di-, tri-, tetra-, and hexanucleotide repeats were found in the plastomes, but no pentanucleotide repeats were detected (Table S4). The total number of SSRs ranged from 55 (*Erysimum sisymbrioides* HM2188) to 136 (*Matthiola stoddartii* HM2157). The number of tandem repeats ranged from 21 (*Lepidium latifolium* HM386) to 108 (*Chorispora sibirica* HM489), and the maximum length of tandem repeats was 272 bp in *Chorispora sibirica* HM489 (Table S4).

In the statistical analysis, plastome length and GC content were used as dependent variables, and maximum length of dispersed repeats, SSR numbers, tandem repeat numbers, and maximum length of tandem repeats were used as independent variables. The results showed that maximum length of dispersed repeats and tandem repeat numbers were positively (coefficient: 1.15×10^{-4} and 4.22×10^{-4}) and significantly (p < 0.05) related to plastome length (Table S5). SSR numbers were negatively (coefficient: -1.44×10^{-4}) and significantly (p < 0.05) related to GC content (Table S5) as most SSRs were A/T repeats.

According to the nucleotide diversity analysis, there were two genes and three intergenic spacers (i.e., *ycf1*, *accD*, *rps15-ycf1*, *rbcL-accD*, and *psbM-trnD*) with higher Pi values, which may serve as effective DNA barcodes for phylogenetic analysis and species identification within Brassicaceae in future studies. In addition, one of the three universal DNA barcodes, *matK*, showed high Pi value; however, the other two universal barcodes, *psbA-trnH* and *rbcL*, had low Pi values (Fig. S3).

Phylogenetic analyses

Alignment length, number of parsimony-informative sites, and GC content of PCGs-con, NPCGs-con, and CP-con were shown in Table 2. Four robust ML trees were reconstructed based on PCGs-con, NPCGs-con, and CP-con. The tree topologies inferred from the three unpartitioned datasets and one partitioned dataset were largely congruent (Figs. 4 and S4-S6). Therefore, only the PCGs-con ML tree is presented and described in the main text (Fig. 4). Combined with the downloaded plastomes, our study covered 24 of the 32 ephemeral species of Brassicaceae. These 24 ephemeral species of Brassicaceae were dispersed across the ML trees (Fig. 4), belonging to 18 genera, i.e., Lepidium (one species), Camelina (one species), Diptychocarpus (one species), Litwinowia (one species), Chorispora (two species), Sterigmostemum (one species), Euclidium (one species), Lachnoloma (one species), Leptaleum (one species), Matthiola (one species), Neotorularia (one species), Tetracme (two species), Strigosella (two species), Alyssum (two species), Meniocus (one species), Goldbachia (two species), Iljinskaea (one species), and Isatis (two species).

Taxa	Voucher	Sample ID	Plastome size (bp)	LSC (bp)	IR (bp)	SSC (bp)	GC content (%)	Number of PCGs (unique)	Number of tRNAs (unique)	Number of rRNAs (unique)	Pseudogene
Alyssum simplex	XJ-92	HM2130	150,682	80,743	26,279	17,381	36.4	85 (78)	37 (30)	8 (4)	rps16
Barbarea vulgaris	Ge130089	HM0089	154,958	83,903	26,504	18,047	36.4	86 (79)	37 (30)	8 (4)	
Camelina microcarpa	XJ-05	HM2043	153,026	82,236	26,484	17,822	36.5	83 (77)	37 (30)	8 (4)	rps16, ycf15
Camelina microcarpa	Q-102	HM1666	153,061	82,236	26,484	17,857	36.5	83 (77)	37 (30)	8 (4)	rps16, ycf15
Capsella orientalis	Ge130613	HM0613	154,549	83,802	26,454	17,839	36.6	86 (79)	37 (30)	8 (4)	
Chorispora sibirica	XJ-120	HM2158	162,326	86,417	32,693	10,523	36.6	86 (78)	37 (30)	8 (4)	accD
Chorispora sibirica	gexj-19287	HM489	162,956	86,590	32,908	10,550	36.5	86 (78)	37 (30)	8 (4)	accD
Dontostemon elegans	XJ-67	HM2105	154,535	83,616	26,446	18,027	35.9	86 (79)	37 (30)	8 (4)	
Dontostemon glandulosus	Ge130056	HM0056	153,628	82,669	26,461	18,037	36.1	86 (79)	37 (30)	8 (4)	
Draba alpina	Q-027	HM1591	152,549	81,993	26,386	17,784	36.5	85 (78)	37 (30)	8 (4)	rps16
Draba alpina	gexj-19293	HM495	152,549	81,993	26,386	17,784	36.5	85 (78)	37 (30)	8 (4)	rps16
Draba fladnizensis	Ge130137	HM0137	153,853	83,037	26,376	18,064	36.5	85 (78)	37 (30)	8 (4)	rps16
Draba melanopus	Ge130049	HM0049	153,799	82,990	26,356	18,097	36.5	85 (78)	37 (30)	8 (4)	rps16
Draba rockii	Ge130131	HM0131	153,930	83,026	26,366	18,172	36.4	85 (78)	37 (30)	8 (4)	rps16
Erysimum flavum	XJ-116	HM2154	154,631	83,817	26,480	17,854	36.6	86 (79)	37 (30)	8 (4)	
Erysimum flavum subsp. altaicum	XJ-152	HM2190	154,580	83,775	26,477	17,851	36.6	86 (79)	37 (30)	8 (4)	
Erysimum hieraciifolium	XJ-156	HM2194	154,581	83,776	26,477	17,851	36.6	86 (79)	37 (30)	8 (4)	
Erysimum quadrangulum	XJ-91	HM2129	154,582	83,777	26,477	17,851	36.6	86 (79)	37 (30)	8 (4)	
Erysimum siliculosum	66-ГХ	HM2137	154,490	83,683	26,477	17,853	36.6	86 (79)	37 (30)	8 (4)	
Erysimum sisymbrioides	XJ-150	HM2188	154,662	83,847	26,479	17,857	36.6	86 (79)	37 (30)	8 (4)	
Euclidium syriacum	XJ-12	HM2050	154,251	83,657	26,466	17,662	36.2	85 (78)	37 (30)	8 (4)	rps16
Eutrema nepalense	Ge130150	HM0150	153,647	83,576	26,062	17,947	36.4	86 (79)	37 (30)	8 (4)	
Goldbachia sabulosa	XJ-76	HM2114	153,454	83,443	26,236	17,539	36.3	86 (79)	37 (30)	8 (4)	
lsatis gymnocarpa	Q-077	HM1641	153,161	83,092	26,230	17,609	36.5	86 (79)	37 (30)	8 (4)	
lsatis gymnocarpa	XJ-130	HM2168	153,168	83,099	26,230	17,609	36.5	86 (79)	37 (30)	8 (4)	
Isatis indigotica	XJ-126	HM2164	153,982	83,766	26,259	17,698	36.5	86 (79)	37 (30)	8 (4)	
Isatis minima	XJ-66	HM2104	153,643	83,424	26,255	17,709	36.5	86 (79)	37 (30)	8 (4)	
Isatis minima	XJ-100	HM2138	153,643	83,423	26,255	17,710	36.5	86 (79)	37 (30)	8 (4)	
Lachnoloma lehmannii	gexj-19117	HM319	153,758	83,273	26,365	17,755	36.2	86 (79)	37 (30)	8 (4)	
Lepidium apetalum	gexj-19096	HM298	154,851	83,896	26,478	17,999	36.4	86 (79)	37 (30)	8 (4)	
Lepidium appelianum	XJ-124	HM2162	153,369	82,380	26,482	18,025	36.5	86 (79)	37 (30)	8 (4)	
Lepidium draba	XJ-117	HM2155	153,797	82,897	26,426	18,048	36.5	86 (79)	37 (30)	8 (4)	
Lepidium ferganense	gexj-19052	HM254	154,725	83,823	26,449	18,004	36.4	86 (79)	37 (30)	8 (4)	
Lepidium latifolium	gexj-19184	HM386	153,804	82,907	26,425	18,047	36.5	86 (79)	37 (30)	8 (4)	

Таха	Voucher	Sample ID	Plastome size (bp)	LSC (bp)	IR (bp)	SSC (bp)	GC content (%)	Number of PCGs (unique)	Number of tRNAs (unique)	Number of rRNAs (unique)	Pseudogene
Lepidium perfoliatum	gexj-19148	HM350	154,386	83,497	26,454	17,981	36.5	86 (79)	37 (30)	8 (4)	
Lepidium sp.	XJ-151	HM2189	153,990	83,561	26,451	17,527	36.5	85 (78)	37 (30)	8 (4)	rps16
Matthiola stoddarti	ХЈ-119	HM2157	153,354	83,345	26,217	17,575	36	86 (79)	37 (30)	8 (4)	
Meniocus linifolius	gexj-19160	HM362	153,411	82,494	26,506	17,905	36.1	85 (78)	37 (30)	8 (4)	rps16
Meniocus linifolius	gexj-19082	HM284	153,375	82,459	26,506	17,904	36.1	85 (78)	37 (30)	8 (4)	rps16
Sisymbrium brassiciforme	TW-0047	HM1247	154,240	84,046	26,250	17,694	36.4	86 (79)	37 (30)	8 (4)	
Sisymbrium loeselii	60-LX	HM2047	154,600	84,429	26,268	17,635	36.4	86 (79)	37 (30)	8 (4)	
Sisymbrium polymorphum	XJ-133	HM2171	154,469	84,298	26,268	17,635	36.4	86 (79)	37 (30)	8 (4)	
Sterigmostemum caspicum	Q-197	HM1761	153,754	83,707	26,250	17,547	36.3	86 (79)	37 (30)	8 (4)	
Strigosella africana	TW-0025	HM1225	154,155	83,141	26,511	17,992	36.3	85 (78)	37 (30)	8 (4)	rps16
Strigosella scorpioides	TW-0010	HM1210	155,018	84,069	26,509	17,931	36.2	86 (79)	37 (30)	8 (4)	
Strigosella scorpioides	gexj-19122	HM324	154,133	83,183	26,509	17,932	36.3	85 (78)	37 (30)	8 (4)	rps16
Strigosella scorpioides	TW-0030	HM1230	154,132	83,181	26,509	17,933	36.3	85 (78)	37 (30)	8 (4)	rps16
Tetracme quadricornis	TW-0094	HM1294	154,488	84,027	26,350	17,761	36.2	86 (79)	37 (30)	8 (4)	
Tetracme recurvata	gexj-19104	HM306	154,562	83,647	26,494	17,927	36.2	86 (79)	37 (30)	8 (4)	

Table 1 (continued)



Fig. 3 Gene maps of newly sequenced plastomes. Only three representatives are shown

 Table 2
 Summary of the three matrices used in maximum likelihood analyses

Alignment name	Alignment length (bp)	Number of variable sites (bp)	Proportion of variable sites (%)	Number of arsimony informative sites (bp)	Proportion of parsimony informative sites (%)	GC content (%)
PCGs-con	51,396	17,795	34.6	11,567	22.5	38.3
NPCGs-con	77,157	44,119	57.2	32,220	41.8	33.5
CP-con	128,553	61,914	48.2	43,787	34.1	35.5

PCGs-con Protein-coding genes concatenated, NPCGs-con Non-protein-coding genes concatenated, CP-con Complete plastomes with one IR removed

Divergence times

The molecular dating analyses in treePL based on PCGscon-div and CP-con-div showed congruent node ages of Brassicaceae (Figs. S7 and S8). For example, the crown age of Brassicaceae was estimated to be 37.73 Mya (95% HPD: 30.96-47.58 Mya) and 36.29 Mya (95% HPD: 35.25-46.25 Mya), and the crown age of core Brassicaceae (i.e., all Brassicaceae excluding tribe Aethionemeae) was 32.70 Mya (95% HPD: 25.86-42.54 Mya) and 32.84 Mya (95% HPD: 30.42-41.42 Mya) in the two treePL analyses, respectively (Table 3). However, the crown ages of Brassicaceae and core Brassicaceae inferred in the MCMCtree analyses (Figs. S9 and S10) were approximately 5 Mya and 1 Mya older than those inferred in the treePL analyses (Table 3). Nonetheless, ephemeral species origination times inferred from treePL and MCMCtree were largely congruent. That is, in the treePL analysis, three ephemeral species occurred in the late Early Miocene, five in the Middle Miocene, 12 in the Late Miocene, and four in the Pliocene and Quaternary, while in the MCMCtree analysis, one ephemeral species originated in the late Early Miocene, four in the Middle Miocene, 15 in the Late Miocene, and four in the Pliocene and Quaternary (Fig. 5; Table S6).

Substitution rate

The PCGs-con ML tree was used to extract substitutions per site (d), and the divergence time (T) was obtained from the stem age of Brassicaceae (46.96 Mya, Fig. S7). Substitutions per site per year for each species was shown in Table S7. The substitution rates of ephemeral were slightly higher than that of the non-ephemeral plants (mean rate: $0.57 \times 10^{-9} > 0.51 \times 10^{-9}$), and t-test showed that the variation was significant (p < 0.01) (Fig. 6).

Phylogenetic signal

Blomberg K was estimated to be 0.084 with a p value of 0.079, and Pagel's λ was 0.245 with a p value of 0.143. The results showed that phylogenetic signals were weak and not significant; therefore, ASR was not performed, and the species origination times (i.e., stem ages of terminal branches) of ephemeral plants were used to represent the independent evolution of ephemeral habit, thus reflecting the evolutionary history of the ephemeral flora (Fig. 5). According to the dated tree, independent evolution of ephemeral habit occurred for at least 20 times (Fig. S7).



Fig. 4 Maximum likelihood tree of Brassicaceae inferred using RAxML based on the PCGs-con dataset. Bootstrap values are shown above branches. Ephemeral plants are colored in red

Discussion

Plastome structural variation and substitution rate variation

In this study, complete plastomes of 49 samples, representing 16 ephemeral and 24 non-ephemeral species, were generated from de novo assembly approach. The observed plastome size in these samples ranged from 150,682 to 162,956 bp, which is within the size range (i.e., 120 to 160 kb) of most land plants [61], and consistent with a previous study on Brassicaceae [11].

IR contraction and expansion are considered important evolutionary events that drive plastid genome size and gene content variations [62, 63]. The IR length of Brassicaceae is relatively conserved at around 26 kb except Chorispora sibirica HM489 and HM2158, which is around 32 kb (Table 1). The dramatic expansion was caused by the presence of double complete *ycf1*, *rps15*, and *ndhH* in the IR regions of *Chorispora sibirica*, but these genes were absent in the IR regions of other plastomes (Fig. 3). In contrast, another species from the same genus, Chorispora tenella (GenBank accession number: NC049622), was only moderately expanded and contained double complete *ycf1* and *rps15* in the IR regions. Although large IR expansions are less common within genus, examples exist in Caryodaphnopsis (20,036-25,601 bp), Euphorbia (26,434-43,573 bp) and Paphiopedilum (31,743-37,043 bp) [64-66], and even within species such as Cinnamomum chartophyllum (20,094-25,974 bp) [9]. IR length variation is intimately connected to double-strand breaks, followed by strand invasion and recombination [67–69], which may be responsible for the dramatic IR expansion in Chorispora.

Chaw and Jansen [70] suggested that the variations in the abundance of smaller repetitive sequences can affect plastome size. In this study, positive and significant correlation was detected between plastome size, maximum length of dispersed repeats and tandem repeat numbers, suggesting that maximum length of dispersed repeats and tandem repeats play an important role in plastome size evolution [9], as has been reported in *Capsicum* [71] and *Medicago* [72]. The SSR and tandem repeat numbers of *Chorispora sibirica* HM489 and HM0613 were higher than those of most other plastomes (Table S4), and some of these repeats may have changed the polarity of the affected segment and gave rise to the inversion of *ycf2*, *ycf15* and *trnL-UUG* [73].

Smith and Donoghue [74] indicated that molecular evolution rates are linked to life history in flowering plants—species with longer generation times have lower substitution rate than species with shorter generation times. Soria-Hernanz et al. [75] indicated that annuals more frequently exhibit faster substitution rates than perennials in *Arabidopsis*, although the underlying

Studies	Crown Brassicaceae (Mya)	Crown core Brassicaceae (Mya)	Dataset	Method
This study	30.96-37.73-47.58	25.86-32.70-42.54	Plastomes (PCGs-con-div)	TreePL
This study	35.25-36.29-46.25	30.42-32.84-41.42	Plastomes (CP-con-div)	TreePL
This study (run1)	31.80-42.22-53.15	24.14-31.36-39.25	Plastomes (PCGs-con-div)	MCMCTree
This study (run2)	31.97-41.81-53.51	22.96-30.78-37.91	Plastomes (PCGs-con-div)	MCMCTree
Hendriks et al. (2023) [52]	25.7-24.5-23.1	22.4-21.1-19.9	297 nuclear genes	TreePL
Hendriks et al. (2023) [52]	29.0-20.2-13.0	24.3-16.9-10.2	Plastomes	TreePL
Huang et al. (2020) [43]	26.8-29.9-33.2	19.6-21.3-22.9	Plastomes	BEAST
Walden et al. (2020) [14]	29.94	25.14	Plastomes	BEAST
Mandáková et al. (2017) [53]	29.4-40.1-54.7	30.6	Plastomes	BEAST
Guo et al. (2017) ^a [12]	30.0-35.2-42.5	21.7-25.3-29.7	Plastomes	MCMCTree
Mohammadin et al. (2017) [54]	37.5-48.0-58.94	35.4	Plastomes	BEAST
Cardinal-McTeague et al. (2016) [55]	40.3-43.4-46.6	38.3	Three plastid loci and two mitochondria loci	BEAST
Huang et al. (2016) ^a [56]	36.3-37.1-37.8	29.1-29.7-30.3	113 nuclear genes	r8s
Hohmann et al. (2015) [13]	27.1-32.4-38.6	19.9-23.4-27.3	Plastomes	BEAST
Edger et al. (2015) [57]	16.8-31.8-45.9	NA	1155 single-copy nuclear genes	BEAST
Couvreur et al. (2010) [58]	24.2-37.6-49.4	20.9-32.3-42.8	Eight nuclear genes, chloroplasts and mitochon- dria	BEAST
Beilstein et al. (2010) [59]	45.2-54.3-64.2	39.4-46.9-54.3	ndhF and PHYA	BEAST
Franzke et al. (2009) [60]	1.0-15.0-35.0	1.0-11.0-28.0	nad4	BEAST

Table 3 Comparison of divergence time estimates with previous studies

^a Time estimation results with the fossil *Thlaspi primaevum* were not shown

mechanism remains unclear [76]. In Brassicaceae, the ephemeral plants complete their life cycle within approximate three month [2], and generally have shorter generation times than non-ephemeral species. In this study, we found faster substitution rates in ephemeral plants than in non-ephemeral plants, which may be due to their different life strategies.

Divergence time within Brassicaceae

Many studies have estimated the divergence times of Brassicaceae using different methods, such as BEAST, MCMCtree or r8s, with various molecular markers, such as ITS, several plastid/nuclear loci, complete plastomes, and hundreds of nuclear genes [12-14, 43, 52-60]. These studies inferred widely varied ages of crown Brassicaceae, ranging from 15.0 to 54.3 Mya (Table 3) [59, 60]; the variation is potentially caused by insufficient parsimony-informative sites in the markers and different fossils used in the dating analyses [56]. In this study, we used plastid coding genes and complete plastomes that contained sufficient parsimony-informative sites to infer the divergence times using TreePL. To compare the influence of the methods, we also performed two parallel analyses using MCMCtree based on plastid coding genes. Although the crown age of Brassicaceae estimated from MCMCtree was approximately 5 Mya older than that from TreePL, the crown age of core Brassicaceae

and origination times of ephemeral plants estimated by the two methods were largely consistent (Tables 3 and S6). Despite the discrepancies in the crown age of Brassicaceae between our study and previous studies (Table 3), it can be concluded that Brassicaceae diversified around the Middle to Late Eocene, and its major clades rapidly originated around or soon after the Eocene-Oligocene transition (EOT) [56, 58].

The origin and evolution of ephemeral flora

Species assembly in the ephemeral flora involves the composition and organization of species within this community, which could be affected by abiotic factors, biotic interactions, species physiological traits, and species evolutionary histories [7]. Therefore, the origin and evolutionary dynamics of the dominant groups in ephemeral flora can be used to infer the evolutionary history of the flora they dwell in. Since ASR cannot be performed due to the weak and non-significant phylogenetic signal [50], we used the origination time of ephemeral plants as proxy to illustrate the evolutionary history of the ephemeral flora. Although our sampling was incomplete at species level-eight ephemeral species and many of their non-ephemeral relatives were not included, some interesting phenomena could nevertheless be found from our dated phylogeny. In addition to the limited sampling ratio, none



Ephemeral species origination time (Mya)

Fig. 5 The origination time of 24 ephemeral species of Brassicaceae. **a** Divergence times were estimated using treePL based on PCGs-con-div; **b** Divergence times were estimated using MCMCtree based on PCGs-con-div. Oligo, Oligocene; Mio, Miocene; Plio, Pliocene; Qua, Quaternary. HPD, highest posterior density



of the 24 ephemeral species included in this study were endemic to Xinjiang. These species exhibited a broad range, including Siberia, Central Asia, or even extending into the Mediterranean region [2]. Therefore, it is likely that the majority of ephemeral plants in Xinjiang were immigrants from other areas, which could bias our understanding of evolutionary history of the ephemeral flora in this region. Nevertheless, the wide current geographic distribution of these species suggests that achieving seed dispersal across long distance were relatively feasible in a short time period. Otherwise, the distant populations would have diverged into distinct species. Consequently, we can infer that the time gap between the origin of ephemeral species and their establishment in Xinjiang was likely quite modest.

According to our estimates, the first occurrence of ephemeral plants in Brassicaceae was in the late Early Miocene (Fig. 5). Mao and Zhang [3] proposed that the ephemeral flora occurred during Pliocene-Pleistocene transition; however, their hypothesis did not undergo a rigorous analysis based on palynological and fossil evidence or a dated phylogeny. Although Li et al. [8] inferred the origination time of ephemeral flora (14-6 Mya) based on a dated phylogeny of Brassicaceae, their results may be biased because there were insufficient parsimonyinformative sites in *trnL-trnF* and ITS which might bias the dating analysis [9, 10]; moreover, they did not consider ephemeral plants of the other families. In this study, we summarized the origination time periods of ephemeral plants from the other families from previous studies (Table 4) [77–91]. According to our study and those previous studies, the origin of ephemeral flora can be dated back to the Early Miocene. For example, Schischkinia, an ephemeral and monotypic genus from Asteraceae, originated at 19.32 Mya [89].

As part of the Central Asia, northern Xinjiang was dominated by widespread aridity during the Oligocene and Early Miocene [92], which was associated with the Paratethys Sea retreat and global cooling since the EOT. In the Oligocene and Early Miocene, conifers and some angiosperms flourished on mountain slopes and river valleys, but little vegetation covered the lowland deserts [93, 94]. Palynological evidence showed that xerophytic herbs remained at low levels in the Junggar Basin/northern Tian Shan [95], which may suggest low species abundance and richness of the ephemeral flora in the Early Miocene. These results were congruent with our findings that only a few ephemeral species from Asteraceae, Brassicaceae and Papaveraceae occurred and occupied the arid lowlands during the Early Miocene (Table 4).

After the Middle Miocene climatic optimum (MMCO, 16.8-14.7 Mya), the global climate became more arid and seasonal, and the atmospheric CO_2 concentration declined [96], which promoted the rise of global dryland flora. For example, Zygophyllaceae, a xerophytic family, rapidly diversified in different continents during the Middle and Late Miocene (15-10 Mya) [97, 98]; and the annual lineages of Astragalus, which are important elements in Central Asian flora, arose in response to progressing aridity during the Late Miocene and Pliocene (8.6-2.98 Mya) [91]. In our study, most Brassicaceae ephemeral species originated in the Middle and Late Miocene, together with the above examples, suggesting that ephemeral flora experienced a rapid species assembly process driven by stepwise intensified aridification during Middle and Late Miocene. In addition to global climate changes, regional tectonic and geological events may also play an important role in the evolution of ephemeral flora. Westerly moisture is the dominant moisture source of Central Asia and has affected herb steppe expansion from the Miocene onwards [93]. Previous studies have suggested that this moisture source has been controlled by the uplift of the Tian Shan and Pamirs Plateau since the Middle Miocene [99, 100]. Around the Miocene-Pliocene transition, the uplift of the Pamir and Central Anatolian Plateau, as well as the collision of the Pamir and Tian Shan ranges, acted as barriers that blocked the eastward transport of water vapor carried by the winter westerly [101, 102]. These events might have led to more intensified seasonality and aridity in Central Asia and created more habitats suitable for the colonization of ephemeral plants from the Pliocene onwards. In such habitats, species that complete their life cycle within one season and spend unfavorable periods as dormant seeds have a high level of fitness [103], which could explain the rapid species assembly of ephemeral flora from different families, such as Apiaceae, Asteraceae

Family	Early Miocene	Middle Miocene	Late Miocene	Pliocene	Pleistocene	Reference
Poaceae				Eremopyrum triticeum		Fan et al. (2009) [77]
			Schismus arabicus			Gallaher et al. (2022) [78]
		Eremopyrum distans				Fan et al. (2013) [79]
Liliaceae			Gagea fragifera	Gagea nigra, G. ova		Peterson et al. (2019) [80]
		Gagea bulbiflora, Tulipa iliensis		Tulipa biflora		Kim and Kim (2018) [81]
Solanaceae			Hyoscyamus pusillus, Physochlaina	capitata, P. physaloides		Tu et al. (2010) [<mark>82</mark>]
Geraniaceae				Erodium hoefftianum		Fiz et al. (2008) [83]
Lamiaceae					Chamaesphacos ilicifolius	Bendiksby (2011) [84]
			Nepeta micrantha, N. pungens, N. f	edtschenkoi		Bendiksby (2011) [84]
Euphoraceae				Euphorbia turczaninowii		Faltner et al. (2023) [85]
Papaveraceae	Roemeria refracta					Peng et al. (2023) [86]
Apiaceae				Ferula sinkiangensis, F. fukanensis, F krylovii, F. dissecta, F. feruloides	: syreitschikowii, F. lehannii, F.	Panahi (2019) [87]
Asteraceae					Tragopogon kasachstanicus, T. elongatus, T. sabulosus, Epilasia acrolasia, E. hemilasia, Scorzonera pusilla, S. sericeolanata, S. circumflexa	Bell et al. (2012) [88]
	Schischkinia albispina					Barres et al. (2013) [89]
Boraginaceae			Nonea caspica	Lappula spinocarpos	Rindera tetraspis, Lappula occultata, L. semiglabra, L. patula, L. duplicicarpa, L. lasiocarpa, Arnebia decumbens	Otero et al. (2019) [90]
Fabaceae			Astragalus arpilobus, A. campylorrh A. commixtus, A. stalinskyi, A. sesarr tribuloides, A. persepolitanus	ynchus, A. bakaliensis, A. vicarious, noides, A. filicaulis, A. oxyglottis, A.		Azani et al. (2019) [91]

Table 4 The origination time of ephemeral plants of Xinjiang from the other families

and Boraginaceae (Table 4), during the Pliocene and Pleistocene.

Future directions

Our study made a comprehensive sampling of ephemeral species of Brassicaceae, and a robust phylogeny was built combined with plastid genomes downloaded from Gen-Bank. Our effort to date the origin of ephemeral species, as well as the evolutionary history of ephemeral flora, was largely accomplished. Despite the efforts, we acknowl-edged that the current sampling ratio of non-ephemeral species was limited, which could bias the occurrence times of ephemeral habit and the historical dynamics of ephemeral flora. Therefore, more species should be sampled in the future study. In addition, nuclear genes should also be used to account for the complex evolutionary relationships between species, genera and tribes, which could bias the evolutionary history of ephemeral flora as well.

Conclusions

In this study, we newly sequenced 49 plastomes of Brassicaceae, representing 16 ephemeral and 24 non-ephemeral species. The plastome comparative analyses showed that *Chorispora sibirica* has an inverted segment (*ycf2*, *ycf15*, trnL-UUG, and their flanking intergenic spacers) near the IR/LSC boundary, and has experienced an extreme IR expansion toward the SSC region, which is caused by the doubled *ycf1*, *rps15*, and *ndhH* in the IR regions. The plastid phylogenomic analyses indicate that ephemeral species are dispersed across the tree of Brassicaceae and have higher molecular evolution rates than the nonephemeral ones. Divergence time estimates showed that non-ephemeral species of Brassicaceae diversified from the Eocene to the present, while ephemeral species occurred in the Early Miocene and mainly diversified during the Middle and Late Miocene. Our findings, together with previous studies, suggest that the ephemeral flora originated in the Early Miocene and experienced relatively rapid species assembly from the Middle Miocene onwards, which may be attributed to paleoclimate changes and regional geological events.

Abbreviations

ASR	Ancestral state reconstruction
BGI	Beijing Genomics Institute
CTAB	Cetyltrimethylammonium bromide
EOT	Eocene-Oligocene transition
IBSC	Herbarium of the South China Botanical Garden of the Chinese
	Academy of Sciences
IR	Inverted repeat
LSC	Large single copy
ML	Maximum likelihood
Муа	Million years ago
NPCG	Non-protein coding gene

PCGProtein coding geneSH-aLRTsShimodaira-Hasegawa-like approximate likelihood ratio testsSRASequence Read ArchiveSSCSmall single copyUFBSUltrafast bootstraps

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-024-04796-0.

Additional file 1: Table S1. Name list of ephemeral species of Brassicaceae from Xinjiang, China. Table S2. Information of the collected samples. Species names, voucher numbers, SRA numbers, and GenBank ID are included. Table S3. The downloaded plastomes and their GenBank accession numbers. Table S4. Repeat analyses results. Number of dispersed repeats, SSRs, and tandem repeats of the 49 newly sequenced plastomes of Brassicaceae are shown. Table S5. The statistics of the relationship between plastome length, GC content, and repeat variables. Table S6. The origination times of the 24 ephemeral species from Brassicaceae. Median and 95% HPD ages from treePL and MCMCtree (run 1) analyses are shown. Table S7. Substitutions per site per year for each species of Brassicaceae. E, ephemeral; No, non-ephemeral.

Additional file 2: Figure S1. Comparison of the SC/IR junctions among the newly generated plastomes of Brassicaceae. Twenty-one were selected as representatives. JLA, LSC/IRa boundary; JSA, SSC/IRa boundary; JSB, SSC/IRb boundary; JLB, LSC/IRb boundary. Figure S2. The mapping results of Chorispora sibirica HM489 and HM2158. Figure S3. The variation of nucleotide diversity across the 49 newly sequenced plastomes. X-axis indicates site positions; y-axis indicates nucleotide diversity. Five hypervariable loci (ycf1, accD, rps15-ycf1, rbcL-accD, and psbM-trnD^{GAC}) and three standard DNA barcodes (psbA-trnH^{GUG}, matK, and rbcL) are indicated. Figure S4. ML tree of Brassicaceae inferred using RAxML based on the NPCGs-con dataset. Bootstrap values are shown above branches. Ephemeral plants are colored in red. Figure S5. ML tree of Brassicaceae inferred using RAxML based on the CP-con dataset. Bootstrap values are shown above branches. Ephemeral plants are colored in red. Figure S6. ML tree of Brassicaceae inferred using IQ-TREE based on the partitioned PCGs-con dataset. Support values of Shimodaira-Hasegawa-like approximate likelihood ratio test (SH-aLRT at the left) and ultrafast bootstrap (UFBS at the right) are shown above the branches, respectively. Figure S7. Divergence time estimation using treePL based on the PCGs-con-div dataset. Numbers near nodes indicate median ages; blue bars indicate 95% HPD. Red stars indicate the origin of ephemeral habit. Figure S8. Divergence time estimation using treePL based on the CP-con-div dataset. Numbers near nodes indicate median ages; blue bars indicate 95% HPD. Figure S9. Divergence time estimation using MCMCtree based on the PCGs-con-div dataset (Parallel run 1). Numbers near nodes indicate median ages; blue bars indicate 95% HPD. Figure S10. Divergence time estimation using MCMCtree based on the PCGs-con-div dataset (Parallel run 2). Numbers near nodes indicate median ages; blue bars indicate 95% HPD.

Acknowledgements

The authors thank Yu-Ying Zhou for DNA extraction, Yan Li for species identification, and Xin-Xin Zhou for providing some of the plant photographs. The authors thank TopEdit (www.topeditsci.com) for its linguistic assistance during the preparation of this manuscript.

Authors' contributions

XJG designed the research. XJG and YF collected the samples. TWX, DQV, and FS performed the analysis. TWX wrote the original draft. TWX, FS, YF, DQV, and XJG reviewed and edited the manuscript. All authors read and approved the final manuscript.

Funding

This study is supported by National Science Foundation of China (No. 32300192, 4227012076, 31770227), Special Foundation for National Science and Technology Basic Research Program of China (2017FY100201).

Availability of data and materials

Raw reads have been submitted to Sequence Read Archive (SRA) under the BioProject PRJNA945952 with SRA accessions SRR23915996–SRR23916044 (Table S2), and China National GeneBank DataBase (CNGBdb) under Project CNP0004145. Plastomes generated in this study have been released in Science Data Bank [104] and GenBank with accessions OQ644446–OQ644494 (Table S2).

Declarations

Ethics approval and consent to participate

Material collections and molecular experiments were performed in compliance with local and national regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 7 July 2023 Accepted: 5 February 2024 Published online: 15 February 2024

References

- Huang G, Li Y, Padilla FM. Ephemeral plants mediate responses of ecosystem carbon exchange to increased precipitation in a temperate desert. Agric For Meteorol. 2015;201:141–52.
- 2. Tang H-P, Yuan S-F. Ephemeral plants in Xinjiang, China. Beijing: High Education Press; 2021.
- 3. Mao Z-M, Zhang D-M. The conspectus of ephemeral flora in northern Xinjiang. Arid Zone Res. 1994;11:1–26.
- Wang X-Q, Jiang J, Lei J-Q, Zhang W-M, Qian Y-B. The distribution of ephemeral plants and their roles in sand stablization in Gurbantunggut Desert. Acta Geogr Sin. 2003;58:598–605.
- Peng M, He H, Wang Z, Li G, Lv X, Pu X, et al. Responses and comprehensive evaluation of growth characteristics of ephemeral plants in the desert–oasis ecotone to soil types. J Environ Manage. 2022;316:115288.
- 6. Losos JB. Phylogenetic perspectives on community ecology. Ecology. 1996;77:1344–54.
- Mucina L. Biome: evolution of a crucial ecological and biogeographical concept. New Phytol. 2019;222:97–114.
- Li Y, Feng Y, Liu B, Lv G-H, Wang X-Y. Phylogenetic relationships and divergence time of Brassicaceae ephemeral plants in Xinjiang. Arid Zone Res. 2014;31:1100–8.
- 9. Xiao T-W, Ge X-J. Plastome structure, phylogenomics, and divergence times of tribe Cinnamomeae (Lauraceae). BMC Genomics. 2022;23:642.
- Brandley MC, Wang Y, Guo X, Montes de Oca AN, Fería-Ortíz M, Hikida T, et al. Accommodating heterogenous rates of evolution in molecular divergence dating methods: an example using intercontinental dispersal of *Plestiodon (Eumeces)* lizards. Syst Biol. 2011;60:3–15.
- Liu LM, Du XY, Guo C, Li DZ. Resolving robust phylogenetic relationships of core Brassicaceae using genome skimming data. J Syst Evol. 2020;59:442–53.
- 12. Guo X, Liu J, Hao G, Zhang L, Mao K, Wang X, et al. Plastome phylogeny and early diversification of Brassicaceae. BMC Genomics. 2017;18:176.
- Hohmann N, Wolf EM, Lysak MA, Koch MA. A time-calibrated road map of Brassicaceae species radiation and evolutionary history. Plant Cell. 2015;27:2770–84.
- Walden N, German DA, Wolf EM, Kiefer M, Rigault P, Huang X-C, et al. Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. Nat Commun. 2020;11:3795.
- Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 1987;19:11–5.
- Liu L, Zhang YQ, Tumi L, Suni ML, Arakaki M, Burgess KS, et al. Genetic markers in Andean *Puya* species (Bromeliaceae) with implications on plastome evolution and phylogeny. Ecol Evol. 2022;12:e9159.

- 17. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90.
- Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:241.
- Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589–95.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9.
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45:W6–11.
- 23. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47:W59–64.
- Amiryousefi A, Hyvönen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34:3030–1.
- 25. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29:4633–42.
- Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33:2583–5.
- 27. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27:573–80.
- R: a language and environment for statistical computing. http://www.Rproject.org/. Accessed 12 Aug 2023.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
- 30. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2.
- 31. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2016.
- 32. Jin J-J. PersonalUtilities. 2019. https://github.com/Kinggerm/PersonalUt ilities. Accessed 15 Oct 2019.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.
- Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 2016;4:e1660.
- Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
- Xi Z, Ruhfel BR, Schaefer H, Amorim AM, Sugumaran M, Wurdack KJ, et al. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. Proc Natl Acad Sci USA. 2012;109:17519–24.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2014;32:268–74.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35:518–22.
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59:307–21.
- 41. FigTree. http://tree.bio.ed.ac.uk/software/figtree/. Accessed 25 Nov 2018.
- 42. Smith SA, O'Meara BC. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics. 2012;28:2689–90.

- Huang X-C, German DA, Koch MA. Temporal patterns of diversification in Brassicaceae demonstrate decoupling of rate shifts and mesopolyploidization events. Ann Bot. 2020;125:29–47.
- Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytol. 2015;207:437–53.
- 45. Franzke A, Koch MA, Mummenhoff K. Turnip time travels: age estimates in Brassicaceae. Trends Plant Sci. 2016;21:554–61.
- Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: a software platform for bayesian evolutionary analysis. PLoS Comp Biol. 2014;10:e1003537.
- Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using Tracer 1.7. Syst Biol. 2018;67:901–4.
- Steenwyk JL, Buida TJ III, Labella AL, Li Y, Shen X-X, Rokas A. PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. Bioinformatics. 2021;37:2325–31.
- Litsios G, Salamin N. Effects of phylogenetic signal on ancestral state reconstruction. Syst Biol. 2012;61:533–8.
- 51. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. 2012;3:217–23.
- Hendriks KP, Kiefer C, Al-Shehbaz IA, Bailey CD, Hooft van Huysduynen A, Nikolov LA, et al. Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset. Curr Biol. 2023;33:4052-4068.e4056.
- Mandáková T, Li Z, Barker MS, Lysak MA. Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. Plant J. 2017;91:3–21.
- Mohammadin S, Peterse K, van de Kerke SJ, Chatrou LW, Dönmez AA, Mummenhoff K, et al. Anatolian origins and diversification of Aethionema, the sister lineage of the core Brassicaceae. Am J Bot. 2017;104:1042–54.
- Cardinal-McTeague WM, Sytsma KJ, Hall JC. Biogeography and diversification of Brassicales: a 103million year tale. Mol Phylogenet Evol. 2016;99:204–24.
- Huang C-H, Sun R, Hu Y, Zeng L, Zhang N, Cai L, et al. Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. Mol Biol Evol. 2016;33:394–412.
- 57. Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Glöckner G, Platts AE, et al. The butterfly plant arms-race escalated by gene and genome duplications. Proc Natl Acad Sci USA. 2015;112:8362–6.
- Couvreur TLP, Franzke A, Al-Shehbaz IA, Bakker FT, Koch MA, Mummenhoff K. Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). Mol Biol Evol. 2010;27:55–71.
- Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S. Dated molecular phylogenies indicate a Miocene origin for *Arabidop-sis thaliana*. Proc Natl Acad Sci USA. 2010;107:18724–8.
- Franzke A, German D, Al-Shehbaz IA, Mummenhoff K. Arabidopsis family ties: molecular phylogeny and age estimates in Brassicaceae. Taxon. 2009;58:425–37.
- Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76:273–97.
- Henriquez CL, Abdullah, Ahmed I, Carlsen MM, Zuluaga A, Croat TB, et al. Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). Genomics. 2020;112:2349–60.
- Abdullah, Henriquez CL, Mehmood F, Carlsen MM, Islam M, Waheed MT, et al. Complete chloroplast genomes of *Anthurium huixtlense* and *Pothos scandens* (Pothoideae, Araceae): unique inverted repeat expansion and contraction affect rate of evolution. J Mol Evol. 2020;88:562–74.
- 64. Wei N, Pérez-Escobar OA, Musili PM, Huang W-C, Yang J-B, Hu A-Q, et al. Plastome evolution in the hyperdiverse genus *Euphorbia* (Euphorbiaceae) using phylogenomic and comparative analyses: large-scale expansion and contraction of the inverted repeat region. Front Plant Sci. 2021;12:712064.
- Guo Y-Y, Yang J-X, Bai M-Z, Zhang G-Q, Liu Z-J. The chloroplast genome evolution of Venus slipper (*Paphiopedilum*): IR expansion, SSC contraction, and highly rearranged SSC regions. BMC Plant Biol. 2021;21:248.

- 66. Song Y, Yu WB, Tan Y, Liu B, Yao X, Jin J, et al. Evolutionary comparisons of the chloroplast genome in Lauraceae and insights into loss events in the Magnoliids. Genome Biol Evol. 2017;9:2354–64.
- Lee C, Choi IS, Cardoso D, de Lima HC, de Queiroz LP, Wojciechowski MF, et al. The chicken or the egg? Plastome evolution and an independent loss of the inverted repeat in papilionoid legumes. Plant J. 2021;107:861–75.
- 68. Goulding SE, Wolfe KH, Olmstead RG, Morden CW. Ebb and flow of the chloroplast inverted repeat. Mol Gen Genet. 1996;252:195–206.
- 69. Wang R-J, Cheng C-L, Chang C-C, Wu C-L, Su T-M, Chaw S-M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. BMC Evol Biol. 2008;8:36.
- Chaw S-M, Jansen RK. Plastid genome evolution, vol. 85. London: Elsevier Academic Press; 2018.
- Jo YD, Park J, Kim J, Song W, Hur C-G, Lee Y-H, et al. Complete sequencing and comparative analyses of the pepper (*Capsicum annuum* L.) plastome revealed high frequency of tandem repeats and large insertion/deletions on pepper plastome. Plant Cell Rep. 2011;30:217–29.
- 72. Wu S, Chen J, Li Y, Liu A, Li A, Yin M, et al. Extensive genomic rearrangements mediated by repetitive sequences in plastomes of *Medicago* and its relatives. BMC Plant Biol. 2021;21:421.
- Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol. 2011;28:583–600.
- 74. Smith SA, Donoghue MJ. Rates of molecular evolution are linked to life history in flowering plants. Science. 2008;322:86–9.
- Soria-Hernanz DF, Fiz-Palacios O, Braverman JM, Hamilton MB. Reconsidering the generation time hypothesis based on nuclear ribosomal ITS sequence comparisons in annual and perennial angiosperms. BMC Evol Biol. 2008;8:344.
- Gaut B, Yang L, Takuno S, Eguiarte LE. The patterns and causes of variation in plant nucleotide substitution rates. Annu Rev Ecol Evol Syst. 2011;42:245–66.
- Fan X, Sha L-N, Yang R-W, Zhang H-Q, Kang H-Y, Ding C-B, et al. Phylogeny and evolutionary history of *Leymus* (Triticeae; Poaceae) based on a single-copy nuclear gene encoding plastid acetyl-CoA carboxylase. BMC Evol Biol. 2009;9:247.
- Gallaher TJ, Peterson PM, Soreng RJ, Zuloaga FO, Li D-Z, Clark LG, et al. Grasses through space and time: an overview of the biogeographical and macroevolutionary history of Poaceae. J Syst Evol. 2022;60:522–69.
- Fan X, Sha L-N, Yu S-B, Wu D-D, Chen X-H, Zhuo X-F, et al. Phylogenetic reconstruction and diversification of the Triticeae (Poaceae) based on single-copy nuclear *Acc1* and *Pgk1* gene data. Biochem Syst Ecol. 2013;50:346–60.
- Peterson A, Harpke D, Peterson J, Harpke A, Peruzzi L. A pre-Miocene Irano-Turanian cradle: origin and diversification of the species-rich monocot genus *Gagea* (Liliaceae). Ecol Evol. 2019;9:5870–90.
- Kim JS, Kim J-H. Updated molecular phylogenetic analysis, dating and biogeographical history of the lily family (Liliaceae: Liliales). Bot J Linn Soc. 2018;187:579–93.
- Tu T, Volis S, Dillon MO, Sun H, Wen J. Dispersals of Hyoscyameae and Mandragoreae (Solanaceae) from the New World to Eurasia in the early Miocene and their biogeographic diversification within Eurasia. Mol Phylogenet Evol. 2010;57:1226–37.
- Fiz O, Vargas P, Alarcón M, Aedo C, García JL, Aldasoro JJ. Phylogeny and historical biogeography of Geraniaceae in relation to climate changes and pollination ecology. Syst Bot. 2008;33:326–42.
- Bendiksby M. Molecular phylogeny, taxonomy, and historical biogeography of Lamiaceae subfamily Lamioideae, including surveys of alloploid speciation in two Eurasian genera, *Galeopsis* and *Lamium*. Oslo: University of Oslo; 2011.
- Faltner F, Wessely J, Frajman B. Phylogenetic data reveal a surprising origin of *Euphorbia orphanidis* (Euphorbiaceae) and environmental modeling suggests that microtopology limits its distribution to small patches in Mt. Parnassus (Greece). Front Plant Sci. 2023;14:1116496.
- Peng H-W, Xiang K-L, Erst AS, Lian L, Ortiz RDC, Jabbour F, et al. A complete genus-level phylogeny reveals the Cretaceous biogeographic diversification of the poppy family. Mol Phylogenet Evol. 2023;181:107712.

- Panahi M. Biogeographic reconstruction of the genus *Ferula* inferred from analyses of nrDNA and cpDNA sequences. Iran J Bot. 2019;25:79–94.
- Bell CD, Mavrodiev EV, Soltis PS, Calaminus AK, Albach DC, Cellinese N, et al. Rapid diversification of *Tragopogon* and ecological associates in Eurasia. J Evol Biol. 2012;25:2470–80.
- Barres L, Sanmartín I, Anderson CL, Susanna A, Buerki S, Galbany-Casals M, et al. Reconstructing the evolution and biogeographic history of tribe Cardueae (Compositae). Am J Bot. 2013;100:867–82.
- Otero A, Jiménez-Mejías P, Valcárcel V, Vargas P. Being in the right place at the right time? Parallel diversification bursts favored by the persistence of ancient epizoochorous traits and hidden factors in Cynoglossoideae. Am J Bot. 2019;106:438–52.
- Azani N, Bruneau A, Wojciechowski MF, Zarre S. Miocene climate change as a driving force for multiple origins of annual species in *Astragalus* (Fabaceae, Papilionoideae). Mol Phylogenet Evol. 2019;137:210–21.
- Guo ZT, Sun B, Zhang ZS, Peng SZ, Xiao GQ, Ge JY, et al. A major reorganization of Asian climate by the early Miocene. Clim Past. 2008;4:153–74.
- Barbolini N, Woutersen A, Dupont-Nivet G, Silvestro D, Tardif D, Coster PMC, et al. Cenozoic evolution of the steppe-desert biome in Central Asia. Sci Adv. 2020;6:eabb8227.
- 94. Wang W-M, Shu J-W. Cenozoic xeromorphic vegetation in China and its spatial and temporal development in connection with global changes. Palaeoworld. 2013;22:86–92.
- Sun J, Zhang Z. Palynological evidence for the Mid-Miocene Climatic Optimum recorded in Cenozoic sediments of the Tian Shan Range, northwestern China. Glob Planet Change. 2008;64:53–68.
- Holbourn A, Kuhnt W, Lyle M, Schneider L, Romero O, Andersen N. Middle Miocene climate cooling linked to intensification of eastern equatorial Pacific upwelling. Geology. 2014;42:19–22.
- 97. Wu S-D, Zhang L-J, Lin L, Yu S-X, Chen Z-D, Wang W. Insights into the historical assembly of global dryland floras: the diversification of Zygo-phyllaceae. BMC Evol Biol. 2018;18:166.
- Wang Q, Wu S, Su X, Zhang L, Xu X, Lyu L, et al. Niche conservatism and elevated diversification shape species diversity in drylands: evidence from Zygophyllaceae. Proc Royal Soc B. 2018;285:20181742.
- Miao Y, Herrmann M, Wu F, Yan X, Yang S. What controlled Mid-Late Miocene long-term aridification in Central Asia? — global cooling or Tibetan Plateau uplift: a review. Earth-Sci Rev. 2012;112:155–72.
- Zhang Z, Sun J. Palynological evidence for Neogene environmental change in the foreland basin of the southern Tianshan range, northwestern China. Glob Planet Change. 2011;75:56–66.
- Meijers MJM, Brocard GY, Cosca MA, Lüdecke T, Teyssier C, Whitney DL, et al. Rapid late Miocene surface uplift of the Central Anatolian Plateau margin. Earth Planet Sci Lett. 2018;497:29–41.
- 102. Thompson JA, Burbank DW, Li T, Chen J, Bookhagen B. Late Miocene northward propagation of the northeast Pamir thrust system, northwest China. Tectonics. 2015;34:510–34.
- Wang J-C, Pan B-R, Albach DC. Evolution of morphological and climatic adaptations in *Veronica* L. (Plantaginaceae). PeerJ. 2016;4:e2333.
- Xiao T-W, Ge X-J. Plastid phylogenomics of ephemeral species of Brassicaceae in Xinjiang, China. Science Data Bank; 2023. https://doi.org/10. 57760/sciencedb.09329.

Publisher's Note

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