



Variation in Plastome Sizes Accompanied by Evolutionary History in Monogenomic Triticeae (Poaceae: Triticeae)

Ning Chen¹, Li-Na Sha¹, Yi-Ling Wang², Ling-Juan Yin³, Yue Zhang¹, Yi Wang¹, Dan-Dan Wu¹, Hou-Yang Kang¹, Hai-Qin Zhang¹, Yong-Hong Zhou¹, Gen-Lou Sun^{4*} and Xing Fan^{1*}

¹ Triticeae Research Institute, Sichuan Agricultural University, Chengdu, China, ² College of Life Science, Shanxi Normal University, Shanxi, China, ³ Lijiang Nationality Secondary Specialized School, Lijiang, China, ⁴ Saint Mary's University, Halifax, NS, Canada

OPEN ACCESS

Edited by:

Peter Poczai,
University of Helsinki, Finland

Reviewed by:

Yingjuan Su,
Sun Yat-sen University, China
Yun-peng Du,
Beijing Academy of Agricultural
and Forestry Sciences, China

*Correspondence:

Gen-Lou Sun
genlou.sun@smu.ca
Xing Fan
fanxing9988@163.com

Specialty section:

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

Received: 14 July 2021

Accepted: 02 November 2021

Published: 13 December 2021

Citation:

Chen N, Sha L-N, Wang Y-L,
Yin L-J, Zhang Y, Wang Y, Wu D-D,
Kang H-Y, Zhang H-Q, Zhou Y-H,
Sun G-L and Fan X (2021) Variation
in Plastome Sizes Accompanied by
Evolutionary History in Monogenomic
Triticeae (Poaceae: Triticeae).
Front. Plant Sci. 12:741063.
doi: 10.3389/fpls.2021.741063

To investigate the pattern of chloroplast genome variation in Triticeae, we comprehensively analyzed the indels in protein-coding genes and intergenic sequence, gene loss/pseudonization, intron variation, expansion/contraction in inverted repeat regions, and the relationship between sequence characteristics and chloroplast genome size in 34 monogenomic Triticeae plants. Ancestral genome reconstruction suggests that major length variations occurred in four-stem branches of monogenomic Triticeae followed by independent changes in each genus. It was shown that the chloroplast genome sizes of monogenomic Triticeae were highly variable. The chloroplast genome of *Pseudoroegneria*, *Dasypyrum*, *Lophopyrum*, *Thinopyrum*, *Eremopyrum*, *Agropyron*, *Australopyrum*, and *Henradia* in Triticeae had evolved toward size reduction largely because of pseudogenes elimination events and length deletion fragments in intergenic. The *Aegilops/Triticum* complex, *Taeniatherum*, *Secale*, *Crithopsis*, *Herteranthelium*, and *Hordeum* in Triticeae had a larger chloroplast genome size. The large size variation in major lineages and their subclades are most likely consequences of adaptive processes since these variations were significantly correlated with divergence time and historical climatic changes. We also found that several intergenic regions, such as *petN-trnC* and *psbE-petL* containing unique genetic information, which can be used as important tools to identify the maternal relationship among Triticeae species. Our results contribute to the novel knowledge of plastid genome evolution in Triticeae.

Keywords: chloroplast genome, Triticeae, genome variation, genome size, IR expansion/contraction

INTRODUCTION

Chloroplast DNA (cp DNA) is composed of a single circular DNA molecule with a quadripartite structure and encodes multiple proteins, including components of light reactions in the photosynthesis process (Martin et al., 2002). In angiosperms, cp genomes have a very constrained size, ranging from 115 to 165 kb in length, and consist of two copies of an inverted repeat

Abbreviations: Cp, chloroplast; Ir, inverted repeat; Lsc, large-single-copy; Ssc, small single copy; Igs, intergenic region sequence; ML, maximum likelihood; Bi, Bayesian inference; Mcmc, Markov Chain Monte Carlo; Bs, bootstrapping.

(IR) region, a large-single-copy (LSC) region, and a small single copy (SSC) region (Raubeson and Jansen, 2005; Wicke et al., 2011). Because of a uniparental mode of inheritance and high conservation in gene content and genome structure, the chloroplast genome is generally treated as a single locus (Raubeson and Jansen, 2005; Nock et al., 2011). With a smaller effective population size and being essentially recombination-free, the chloroplast genome has a shorter coalescent time than nuclear genomes (Birky et al., 1983). The cp genome is highly conserved in structure, size, and gene content within land plants (Palmer, 1985; Wicke et al., 2011), because chloroplast sequences evolve at approximately half the speed of nuclear regions (Walker et al., 2014). However, variations in some regions of the cp genome have been widely reported in plants (Krause, 2011; Weng et al., 2014; Schwarz et al., 2015; Chen et al., 2017; Bedoya et al., 2019; Shrestha et al., 2019). These advantageous features give the chloroplast genome an important value in reconstructing phylogeny, DNA barcoding for accurate identification of plant species, and tracing evolutionary history (Jansen et al., 2008; Walker et al., 2014; Chen et al., 2017; Bedoya et al., 2019; Shrestha et al., 2019). However, associations between DNA composition and cp genome divergence need to be clarified in species over a range of evolutionary time.

Comparison of whole cp genomes can explore sequence variation, permit examination of molecular evolutionary patterns associated with structural rearrangement, and elucidate genetic changes underlying those events. Previous studies on seed plants have suggested that structural rearrangement, intergenic region variation, variation in IR regions, and gene loss were principal factors driving the variation in chloroplast genome size and structure (Kugita et al., 2003; Parks et al., 2009; Wu et al., 2011; Chen et al., 2017; Zheng et al., 2017). It has been reported that contractions into single copy regions with inversions have shaped certain evolutionary features for monophyletic groups (Palmer et al., 1985; Hoot and Palmer, 1994). While these described features in chloroplast genome variation enable researchers to investigate genome divergences over a broad range of evolutionary time, from early land plants to recently domesticated plants, comparisons among very distant relatives have yielded results with uncertain generality (Matsuoka et al., 2002; Zheng et al., 2017). Moreover, a comprehensive search for factors that drive the variation in genome size in a given phylogenetic framework is still lacking.

The wheat tribe (*Poaceae*: Triticeae), an economically important gene pool for genetic improvement of cereal and forage crops, includes about 450 diploid and polyploid species that distribute in a wide range of ecological habitats over temperate, subtropical, and tropic alpine regions (Dewey, 1984). In Triticeae, 24 major basic genomes in diploid species have been designated and recognized in 18 monogenomic genera (Löve, 1984; Wang et al., 1994). Monogenomic genera have been the focus of numerous evolutionary investigations, partly because of either their economic importance (including barley, rye, and crested-wheat grasses) or their genome donor to the speciation of polyploid species (accounting for 75% of the Triticeae species), as well as their wide variety of species richness, morphology, ecology, and distribution. Previous data from geographical

distribution (Sakamoto, 1973) and morphological characteristics (West et al., 1988), and DNA information (Kellogg et al., 1996) have suggested that the first step of phylogenetic diversification in Triticeae occurred at the diploid level and then gave rise to different present-day diploid lineages. Since the tribe was originated, lineage diversification of monogenomic genus in a range of evolutionary time raised the question of how cp genomes have evolved. Comparison of cp genomic structure on several monogenomic Triticeae showed that a 737-bp deletion in *Pseudoroegneria libanotica* might be related to potential nuclear-cytoplasm transfer (Wu et al., 2017). Analysis of the cp genome sequence in *Agropyron cristatum* suggested that deletion of *accD* and translocation of *rpl23* genes might represent an independent gene-loss event or an additional divergence in Triticeae (Chen et al., 2017). However, to better understand the pattern of variation in cp genome structure and size in Triticeae, it is critical to assess the evolutionary processes that drive cp genome variation accompanied by the diversification of Triticeae with relatively well-sampled monogenomic Triticeae genera.

Here, we carried out phylogenetic reconstructions and estimated diversification patterns using the cp genome from 34 species of 17 monogenomic genera within the Triticeae. While generic definitions within the Triticeae have been extremely variable (Barkworth, 1998), our sample represents nearly all monogenomic genera and genome types according to genome-based classifications of the tribe (Yen et al., 2005). The aims of this study were to (1) examine variation in cp genomic structure and size in monogenomic Triticeae; (2) document the process of cp genomic variation accompanied by the diversification history of monogenomic Triticeae. Knowledge of cp genomic variation over a range of evolutionary time would provide a better understanding of the evolutionary history of the Triticeae.

MATERIALS AND METHODS

Data Collection

Thirty-five species, which included 34 individuals representing 17 genera and 20 basic genomes within Triticeae, were sampled (Table 1). *Brachypodium distachyon* was included as an outgroup. Twenty chloroplast genome sequences are highlighted in bold in Table 1, which are from *Triticum-Aegilops* complex, *Hordeum bulbosum*, *Hordeum vulgare*, *Hordeum vulgare* ssp. *spontaneum*, *Secale cereale*, were downloaded from the NCBI published data (Saski et al., 2006; Gogniashvili et al., 2014, NCBI web, Bortiri et al., 2008; Gornicki et al., 2014; Middleton et al., 2014; Saarela et al., 2015). The remaining sequences were from our previous publications (Chen et al., 2017, 2020; Wu et al., 2017).

Comparison of 35 Chloroplast Genomes

The software mVISTA was used to compare the cp genomes using the annotation of *A. cristatum* as a reference (Frazer et al., 2004). The sequences were aligned with MAFFT v6.833 (Katoh and Toh, 2010) using default settings. Sequence variation, such as gene loss, divergent genes, intergenic sequence (IGS) indels, and IR contraction/expansion, were examined by MEGA 6.0

TABLE 1 | List of taxa used in this study.

No.	Sequence number	Species	Accession no.	Genome	Ploidy	Origin	GeneBank	Chloroplast genome size (bp)
1	KJ614418	<i>Aegilops bicornis</i>	Clae57	S ^b	2×	SE Mediterranean	KJ614418	136,861
2	KJ614416	<i>Aegilops longissima</i>	TA1924	S ^l	2×	E Mediterranean	KJ614416	136,875
3	KY636033	<i>Aegilops markgrafii</i>	AE 1381	C	2×	NE Mediterranean	KY636033	136,063
4	KJ614413	<i>Aegilops searsii</i>	TA1926	S ^S	2×	E Mediterranean	KJ614413	136,870
5	KJ614419	<i>Aegilops sharonensis</i>	TA1995	S ^{Sh}	2×	Israel, Lebanon	KJ614419	136,867
6	KJ614406	<i>Aegilops speltoides</i>	PI487232	S	2×	E Mediterranean	KJ614406	135,652
7	KJ614405	<i>Aegilops speltoides</i> ssp. <i>ligustica</i>	TA1796	S	2×	E Mediterranean	KJ614405	135,660
8	KJ614412	<i>Aegilops tauschii</i>	Cultivar AL8/78	D	2×	SW-C Asia	KJ614412	135,568
10	KY636059	<i>Aegilops umbellulata</i>	AE 153	U	2×	SE Europe, SW Asia	KY636059	136,028
9	KY636056	<i>Aegilops umbellulata</i> ssp. <i>transcaucasica</i>	AE 1070	U	2×	SE Europe, SW Asia	KY636056	136,031
11	C22	<i>Agropyron cristatum</i>	PI 598628	P	2×	Kazakhstan	KY126307	135,554
12	K30	<i>Agropyron mongolicum</i>	PI 499392	P	2×	Nei Monggol, China	MH285848	135,547
13	KY636075	<i>Amblyopyrum muticum</i>	PI 560125	T	2×	Turkey	KY636075	135,787
14	E48	<i>Australopyrum retrofractum</i>	PI 533013	W	2×	NEW South Wales, Australia	MH331642	135,417
15	CRK28	<i>Crithopsis delileana</i>	ND	K	2×	Greece	MH285849	136,436
16	DAK25	<i>Dasyphyrum villosum</i>	W6 7264	V	2×	Greece	MH285850	135,249
17	ERK39	<i>Eremopyrum distans</i>	TA 2229	F	2×	Afghanistan	MH285851	135,589
18	ERK31	<i>Eremopyrum tririceum</i>	Y206	Xe	2×	Xinjiang, China	MH285852	135,554
19	HEK26	<i>Henradia persica</i>	–	O	2×	Iran	MH285853	135,659
20	HEK24	<i>Heterantherium piliferum</i>	PI401352	Q	2×	Iran	MH285854	136,768
21	E46	<i>Hordeum bogdanii</i>	PI 531761	H	2×	Xinjiang, China	MH331641	136,968
22	KM974741	<i>Hordeum jubatum</i>	CAN:Saarela 18478	H	3×	–	KM974741	136,826
23	EF115541	<i>Hordeum vulgare</i>	Morex	I	2×	Minnesota, United States	EF115541	136,462
24	KC912689	<i>Hordeum vulgare</i> ssp. <i>spontaneum</i>	FT462	I	2×	Minnesota, United States	KC912689	136,043
25	K27	<i>Lophopyrum elongatum</i>	PI531718	E ^o	2×	Tunisia	MH331643	135,020
26	C26	<i>Psathyrostachys juncea</i>	PI 430871	Ns	2×	Former soviet union	MH331640	136,597
27	E47	<i>Pseudoroegneria libanotica</i>	PI 228392	St	2×	Iran	KX822019	135,026
28	PSK29	<i>Pseudoroegneria spicata</i>	PI 232134	St	2×	Wyoming, United States	MH285855	135,165
29	KC912691	<i>Secale cereale</i>	Imperial	R	2×	Turkey	KC912691	135,564
30	TAK23	<i>Taeniatherum</i> <i>caput-medusae</i>	PI220591	Ta	2×	Afghanistan	MH285856	136,861
31	THC25	<i>Thinopyrum bessarabicum</i>	PI 531712	E ^b	2×	Estonia	MH331639	135,003
32	LC005977	<i>Triticum monococcum</i>	–	A	2×	Turkey	LC005977	136,886
33	KC912692	<i>Triticum monococcum</i> ssp. <i>aegilopoides</i>	–	A	2×	Turkey	KC912692	136,870
34	KJ614411	<i>Triticum urartu</i>	PI428335	A	2×	E Mediterranean, Caucasus	KJ614411	136,865
35	EU325680	<i>Brachypodium distachyon</i>	Cultivar Bd21	–	–	–	EU325680	135,199

The GenBank accession number with bold represent previously published sequences from the GenBank (<http://www.ncbi.nlm.nih.gov>). ND: not determined.

(Tamura et al., 2013). The IR expansion and contraction of cp genomes were analyzed using IRscope (Amiryousefi et al., 2018).

Phylogenetic Analysis

Phylogenetic analysis was performed using maximum likelihood (ML) and Bayesian inference (BI). ML analysis was conducted using RAxML v.8.2.104 (Stamatakis, 2014). The best model was determined by the model test-ng-0.1.6 software with a default parameter (Kozlov et al., 2019). The optimal model identified was GTR + I + G, which was used in both the ML and BI analyses. The robustness of the trees was estimated by bootstrap support (BS). Statistical support for nodes in ML analysis was estimated using 1,000 fast bootstrap replicates, each with three replicates of stepwise random taxon addition. A BS value of less than 50% was not included in the figures. The BI analysis was performed using MrBayes v3.0 (Huelsenbeck and Ronquist, 2001). Four Markov Chain Monte Carlo (MCMC) chains (one cold and three heated), with MrBayes default heating values ($t = 0.2$), were run for 100,000 generations, each sampling every 100 generations. The first 250 trees were stationarily discarded as “burn-in.” The statistical confidence in nodes was estimated by posterior probabilities (PP). A PP value of less than 0.9 was not included in the figures.

Ancestral State Reconstruction

Two data matrixes, complete cp genome sequence data and concatenated non-protein coding data, were used to trace the ancestral states of genome size on the phylogenetic tree using weighted squared-change parsimony in the software Mesquite v2.5 (Maddison and Maddison, 2021). Weighted squared-change parsimony minimizes the sum of squared change along all branches of the tree, weighting branches by their length (Finarelli and Flynn, 2006). The evolutionary history of the two selected data characters was mapped over the single best ML tree based on complete cp genome sequences.

Divergence Dating

Divergence times with 95% confidence intervals were estimated using the Bayesian relaxed molecular clock method, implemented in BEAST v1.4.6 (Drummond and Rambaut, 2007). Calibration points were performed using a relaxed uncorrelated lognormal molecular clock. A complete cp genome sequence dataset was used to conduct the BEAST analysis. The lack of fossils for Triticeae precluded the direct calibration of tree topologies. Instead, dating was based on the divergence time for the basal-most split in Triticeae (Marcussen et al., 2014). Priors on Triticeae crown age ($15.32 \text{ Ma} \pm 0.34$) were set as inferred by Marcussen et al. (2014), where several macrofossils from grass (*Festuca*, *Berriochloa*, and *Nassella*) were used to calibrate the age of Triticeae. Tracer v1.4 (Rambaut and Drummond, 2018) was used to ensure the convergence of mixing in terms of effective sample size (ESS) values and coefficient rate. The results will be accepted if the values of the estimated sample size were larger than 200, suggesting little autocorrelation between samples. The resultant trees were analyzed using TreeAnnotator in BEAST where the “burn-in” (2,000 trees) was removed, and a maximum credibility

tree was constructed. The trees were then viewed in FigTree v. 1.3.1.¹

Statistical Analysis

Statistical analysis was performed by using the R software. Data statistics of the complete cp genome, protein coding region, and non-protein coding region of each clade were compared with Boxplot. Kruskal–Wallis test was performed on the complete cp genome, protein coding region, and non-protein coding region to determine significant differences among the clades. The relationship between divergence time and indel amount was evaluated by using Spearman’s correlation coefficient test.

RESULTS

Phylogeny of the Triticeae Species

Chloroplast genomes contain an abundance of phylogenetic information, which has been widely used for phylogeny reconstruction at different taxonomic levels, such as order, family, genus, and species, in plants. Using chloroplast genome data, long-standing controversies related to various phylogenetically difficult groups have been resolved, supporting its importance in systematic studies. To better determine the phylogenetic position of Triticeae and further clarify the evolutionary relationships within the Triticeae tribe, phylogenetic analyses were constructed based on the 34 complete chloroplast genomes using *B. distachyon* as an outgroup. Phylogenetic reconstruction based on complete cp genome data resulted in a tree with high posterior probability support across most clades (Figure 1A). The ML and Bayesian analyses of complete cp genome data generated the same tree topology with BS >50% above and PP >0.9 below branches (Figure 1A). The tree illustrated that four clades (I–IV) were recognized. Clade I included the *Aegilops/Triticum* complex, *Taeniatherum* (Ta), *Secale* (R), *Crithopsis* (K), and *Herterantheium* (Q), all of which are members of Mediterranean Triticeae. Clade II contained *Pseudoroegneria* (St), *Dasypyrum* (V), *Lophopyrum* (E^c), and *Thinopyrum* (E^b). Clade III consisted of *Eremopyrum* (F and Xe), *Agropyron* (P), *Australopyrum* (W), and *Henradia* (O). Clade IV comprised the *Hordeum* (H and I) species.

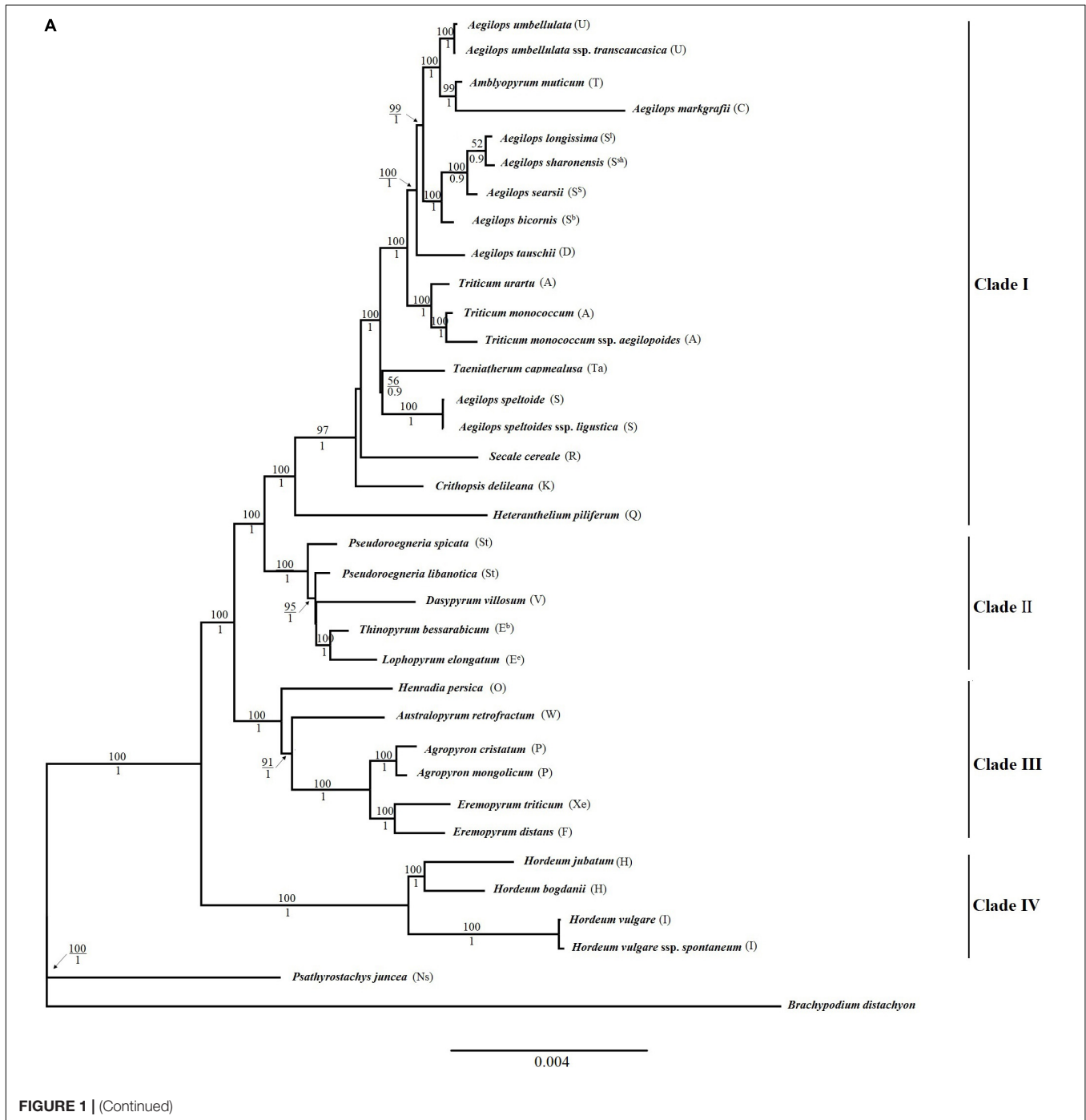
General Variation of Chloroplast Genomes in Triticeae Species

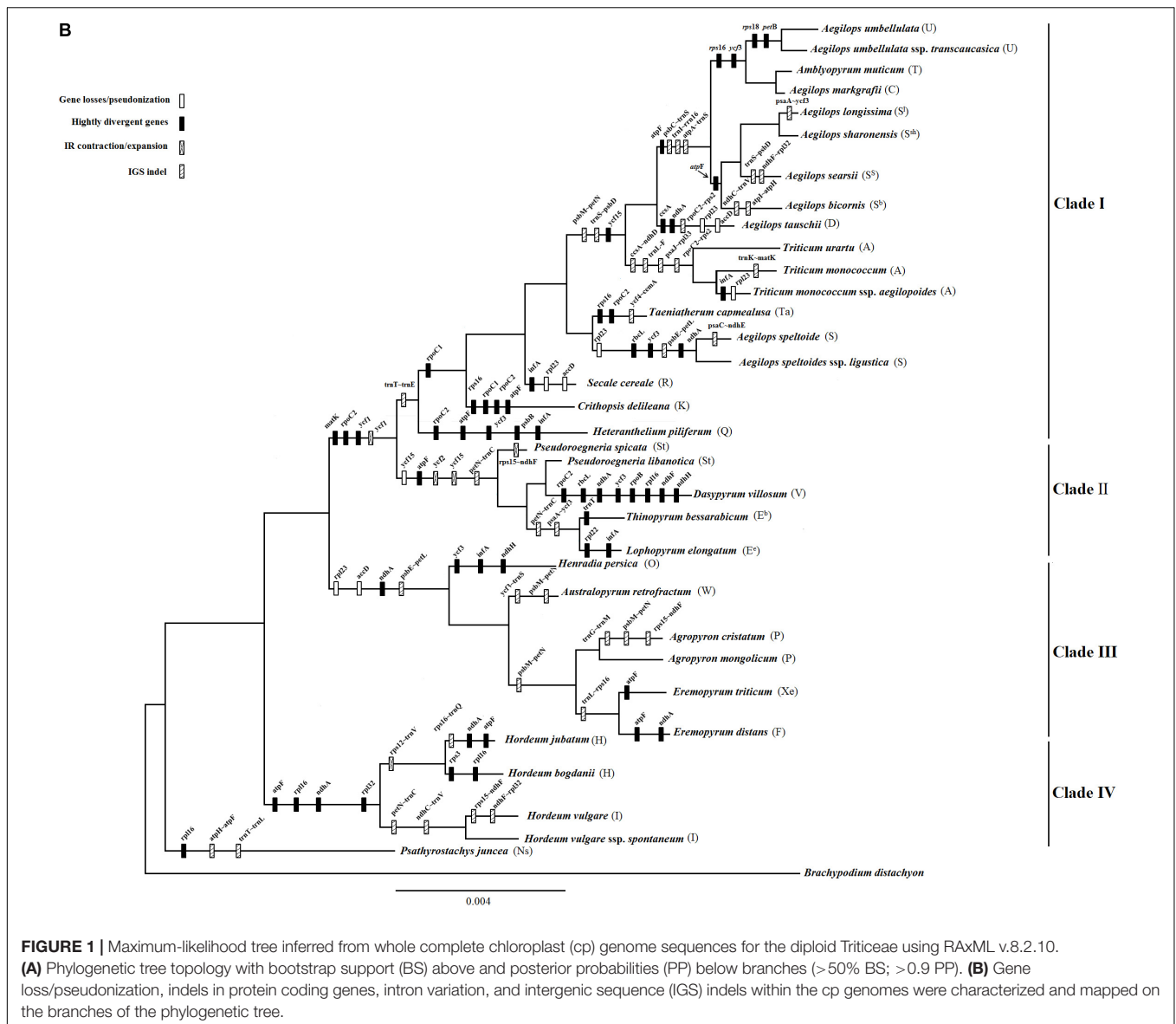
The complete genome size of 34 cp genomes in Triticeae ranged from 135,003 (*Thinopyrum bessarabicum*) to 136,968 bp (*Hordeum bogdanii*), and non-protein coding region size from 75,694 (*T. bessarabicum*) to 77,672 bp (*H. bogdanii*). The level of sequence divergence among the 34 cp genomes was compared using the mVISTA program. We found that non-protein coding regions were more highly variable than protein coding regions, and IRs had lower sequence divergence than SC regions (Supplementary Figure 1). Otherwise, large hotspot variation regions were detected in *petN-rpoB*, *rbcl-psaI*, and *rpl23-ndhB*. Gene loss/pseudonization, indels in protein coding

¹<http://tree.bio.ed.ac.uk/>

genes, intron variation, and IGS indels within the cp genomes were characterized and mapped on branches of the phylogenetic tree (Figure 1A) of the Triticeae species based on complete plastid genomes (Figure 1B). One of the mutation events occurred between the *rbcL* gene and the *psaI* gene in the LSC region (Supplementary Figure 2). This region primarily contains the *rpl23* gene, the *accD* gene, and intergenic regions. The *rpl23* and *accD* genes were completely absent in *A. cristatum*, *Agropyron mongolicum*, *Ere. triticum*, *Ere. distant*, *Australopyrum*

retrofractum, *Henradia persica*, and *Aegilops tauschii*. Within this region, the absence of a long IGS and the *accD* gene was also detected in *Aegilops speltoides* and *Triticum monococcum* ssp. *aegilopoides* (Supplementary Figure 3). Another gene loss event was identified between the *trnL* gene and the *trnI* gene in the IRs region (Supplementary Figure 4). These deletion regions primarily contain a *ycf2* gene fragment, the *ycf15* gene, and intergenic regions, and occurred in *Pseudoroegneria spicata*, *P. libanotica*, *T. bessarabicum*, *Lophopyrum elongatum*, and





Dasyphyrum villosum, which were grouped in clade II (Figure 1). A hot variation region was also found in the non-protein coding regions. All species in Clade II had a 529-bp deletion between the *petN* and *trnC* genes in LSC (Supplementary Figure 5). All species in clade III, *A. speltoides*, *Aegilops speltoides* ssp. *ligustica* and *T. monococcum* ssp. *aegilopoides*, had a 438-bp deletion between *psbE* and *petL* in LSC (Supplementary Figure 6). All species in Clade I and *Hordeum jubatum* contained a 172-bp deletion between *trnT* and *trnE* in LSC.

Highly Divergent Genes

The total number of protein coding genes was 76 in each cp genome of the Triticeae species. Nucleotide divergence occurred in the coding regions of *rpoC2*, *rps3*, *rpl22*, *matK*, *ycf1*, *rps16*, *rpoC1*, *ndhH*, *atpF*, *rpl16*, *rpl32*, *ndhA*, *psbB*, *ycf3*, and *infA*. Among these, the *rpoC2* gene has a 6-bp insertion fragment in the

species clustered in clades I and II, and a 42-bp insertion fragment in *D. villosum* (Supplementary Figures 7, 8). The *Rps3* gene in *H. bogdani* has a 45-bp deletion (Supplementary Figure 9). The *InfA* gene had a deletion of two different 18-bp fragments in *L. elongatum*, *S. cereale*, and *H. jubatum* (Supplementary Figures 10, 11). A 15-bp deletion in the *rpl22* gene was found in *L. elongatum* (Supplementary Figure 12).

Inverted Repeat Contraction/Expansion

The most common events underlying changes in the plastome size of land plants included the contraction/expansion of IR. Land plants have a highly conserved chloroplast genome, but four junctions (LSC/IRB/SSC/IRA) vary in genome size and could affect IR contraction and expansion (Saski et al., 2007; Choi and Park, 2015). All IR boundaries among the sampled monogenomic Triticeae species were in a similar location, but IR size varied from

20,806 (*D. villosum*) to 21,589 bp (*Eremopyrum distans*) because of variation in the size of pseudogenes *ycf2* and *ycf15*. Because of a ~800-bp deletion of *ycf2* and *ycf15* in the clade II species, IR size varied from 20,806 (*D. villosum*) to 20,866 bp (*P. spicata*), which was much smaller than that in the other Triticeae species (>21 kb) (**Figure 2**).

The LSC/IRb/SSC/IRa boundary regions were compared using IRscope (**Figure 2**). The LSC/IRb junction was between the *rpl22* gene and the *rps19* gene in 31 Triticeae species, and close to the *rps19* gene in the plastomes of *Aegilops markgrafii*, *Aegilops umbellulata*, *Aegilops umbellulata* ssp. *transcaucasica*, *Amblyopyrum muticum*, *D. villosum*, *H. jubatum*, and *T. monococcum* ssp. *aegilopoides*, varying from 17 to 36 bp. The *rpl22* gene in the LSC region was close to the LSC/IRb junction in 24 out of the 31 cp genomes and varied from 28 to 42 bp apart from the LSC/IRb junction. The *Rpl2* gene in *H. vulgare* was in IR regions 54 bp away from the LSC/IRb junction, while the *rpl2* gene in *H. vulgare* ssp. *spontaneum* was in the LSC region 5 bp away from the LSC/IRb junction. The distance between *rps15* and the junction of the IRb/SSC region ranged from 13 to 480 bp (except *rps15* in *H. vulgare* that crossed the junction of the IRb/SSC region). At the junction of the SSC/IRa region, the *ndhH* gene crossed the SSC/IRa boundary in all the genomes with 0–1,007 bp located in the SSC region. The *rps19* gene was in the IRa region in 31 of the species 1–51 bp apart from the LSC/IRa junction. The *rpl2* gene was located in the IRa region 590 bp apart from the LSC/IRa junction in *H. vulgare*, and in the LSC region 4 bp apart from the junction in *H. vulgare* ssp. *spontaneum*.

Ancestral State Reconstruction

Sizes of complete chloroplast genomes, protein coding sequences, and non-protein coding sequences in monogenomic Triticeae are shown in **Table 2**. The reconstruction of character evolution revealed that the sizes of complete cp genome sequences and non-protein coding sequences of the species within clades I and IV were gradually enlarged (**Figure 3**). The sizes of complete genome sequences ranged from 135,564 (*S. cereale*) to 136,886 bp (*Triticum monococcum*), and from 136,043 (*H. vulgare* ssp. *spontaneum*) to 136,968 bp (*H. bogdanii*) for the species in clades I and IV, respectively, while the sizes of complete cp genome sequences of species within clades II and III were significantly smaller than those within clades I and IV, and ranged from 135,003 bp (*T. bessarabicum*) to 135,249 bp (*D. villosum*) in clade II, and from 135,417 (*A. retrofractum*) to 135,659 bp (*H. persica*) in clade III. The sizes of protein coding sequences varied from 58,302 (*A. muticum*) to 61,049 bp (*T. monococcum* ssp. *aegilopoides*); from 59,098 (*H. vulgare* ssp. *spontaneum*) to 59,360 bp (*H. jubatum*); from 59,309 (*T. bessarabicum*) to 59,372 bp (*D. villosum*); and from 59,271 (*Eremopyrum tririceum*) to 59,341 bp (*E. distans*) for the species within clades I, IV, II, and III, respectively.

Divergence Dating

Based on the complete cp genome sequences dataset, divergence times with 95% CI by using BEAST analyses generated a maternal time-calibrated tree (**Figure 4**). The divergence time was marked

on 33 branch nodes within the ML tree. Time calibration analysis demonstrated that the time to the most recent common maternal ancestor of the Triticeae was dated to 27.63 MYA (95% CI). The maternal ancestor of *Hordeum* originated about 15.09 MYA (95% CI). The divergence time of maternal ancestor of species in clade III (*Agropyron*, *Australopyrum*, *Henradia*, and *Eremopyrum*) and clade II (*Dasyphyrum*, *Pseudoroegneria*, *Lophopyrum*, and *Thinopyrum*) was dated to 17.16 MYA (95% CI) and 12.62 MYA (95% CI), respectively. The divergence time of the maternal ancestor of *Aegilops/Triticum*, *Taeniatherum*, *Secale*, *Crithopsis*, and *Heteranthelium* in clade I was 12.71 MYA (95% CI).

Statistical Analysis

The insertion and deletion of 33 divergence nodes of the phylogenomic tree are counted in **Table 3**. The divergence time was highly correlated with the number of variations ($R = 0.87$, $p = 1.6e^{-10} < 0.05$) (**Figure 5A**), as well as cp genome size and indel number ($R = 0.37$, $p = 0.04$) (**Figure 5B**). Therefore, both early and late divergence times might lead to a similar trend in the associated number of indels. The highest frequency of indels occurred about 5 MYA.

The sizes of complete genome sequences and non-protein coding sequences of the species within clades I (*Triticum/Aegilops*) and IV (*Hordeum*) were significantly larger than those of the species within clades II (St/E/V genomes) and III (P/F/Xe/W/O genomes) (**Figures 6A,C**). However, the sizes of the protein coding sequences among these species were not significantly different (**Figure 6B**). The average sizes of complete cp genome sequences of the species within clades I–IV were 136,362, 135,093, 135,553, and 136,575 bp (**Figure 6A**), and protein-coding gene sequences were 59,308, 59,324, 59,305, and 59,266 bp (**Figure 6B**), respectively.

DISCUSSION

Plastome size variations within major lineages are mainly attributable to a combination of three factors: differences in length of intergenic regions, changes in intron content, and contraction/expansion of IRs, which resulted in cp genome size ranging from 115 to 165 kb in the length of angiosperms (Raubeson and Jansen, 2005; Brouard et al., 2010; Wicke et al., 2011; Smith et al., 2013; Turmel et al., 2017). Previous studies have shown that the size of the cp genome of Triticeae species ranged from 134 to 137 kb (Gornicki et al., 2014; Middleton et al., 2014; Chen et al., 2017). The intergenic regions, which represent up to 68% of the genome, contribute to most of the observed genome size variation (Turmel et al., 2015). Our ancestral genome reconstruction suggests that major rearrangements occurred in four branches (clades I–IV) of monogenomic Triticeae followed by minor independent rearrangements in each genus. The sizes of complete genome sequences of the species within clades I (*Triticum/Aegilops* complex, *Taeniatherum*, *Secale*, *Crithopsis*, and *Herteranthelium*) and IV (*Hordeum*) were significantly larger than those of within clades II (*Pseudoroegneria*, *Dasyphyrum*, *Lophopyrum*, and

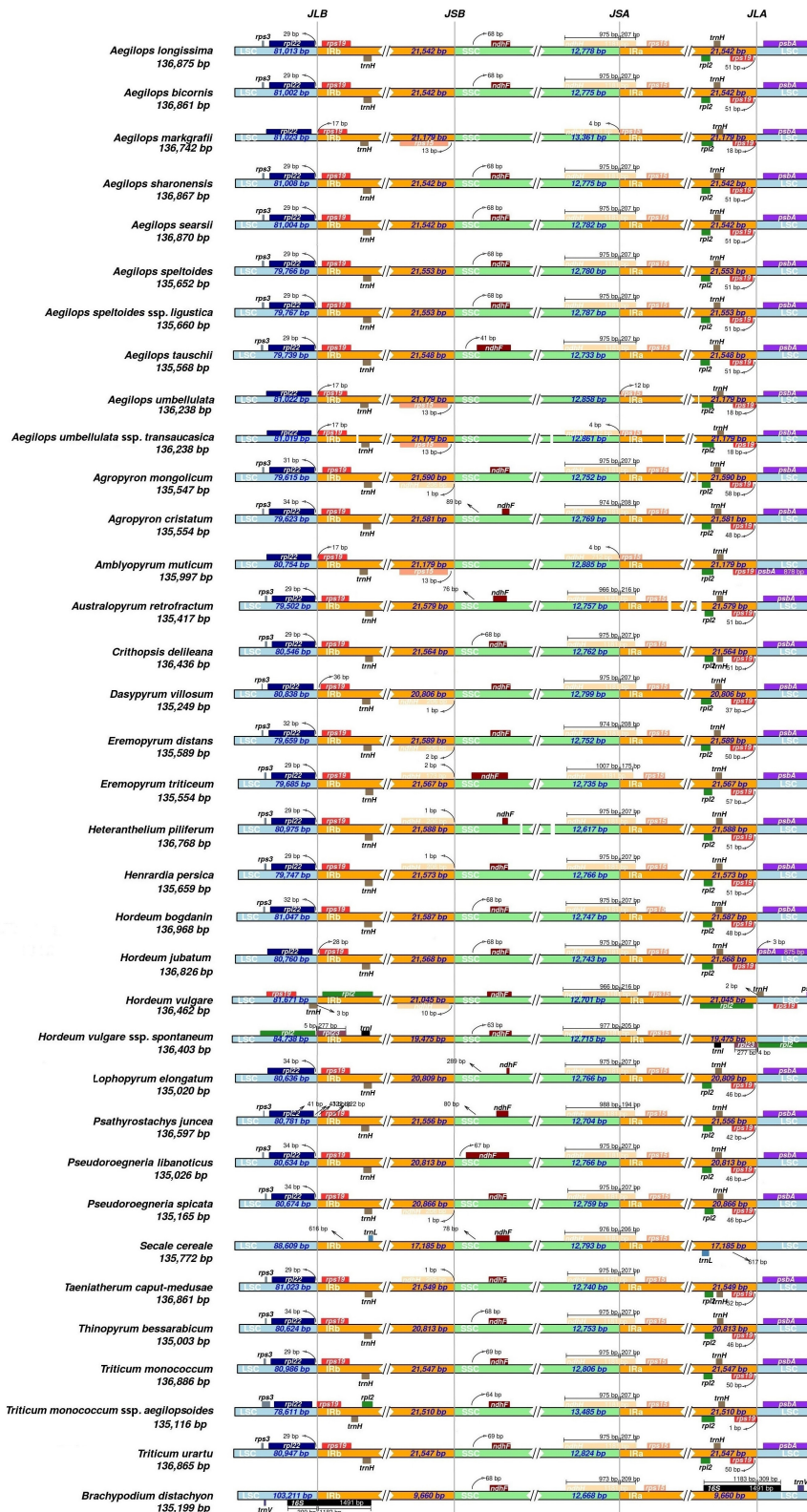


FIGURE 2 | Comparison of the borders of the large-single-copy (LSC) (blue), small single copy (SSC) (green), and inverted repeat (IR) (orange) regions among the 34 cp genomes.

TABLE 2 | Sizes of 34 Triticeae species with complete chloroplast genomes and non-protein/protein coding sequences in four clades.

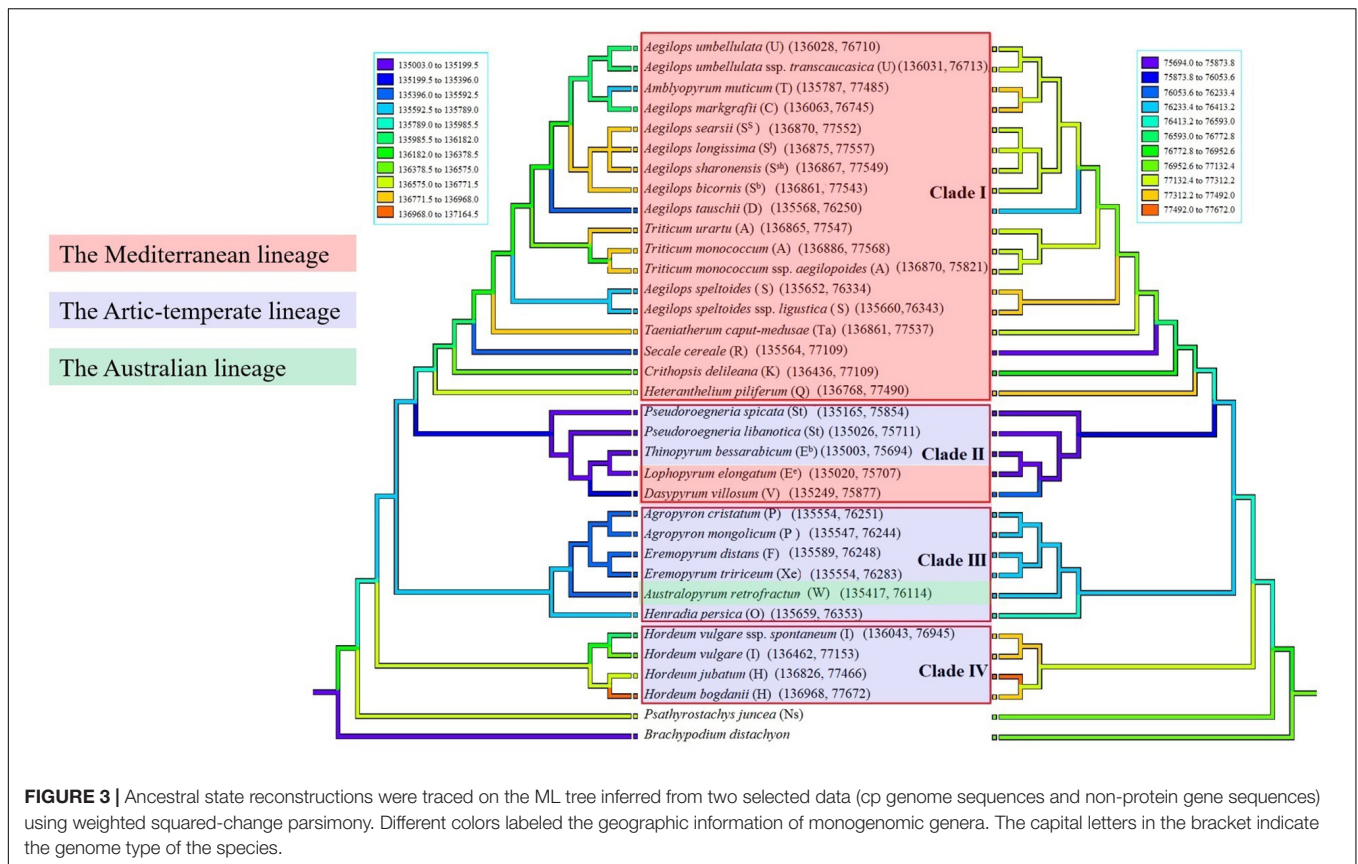
No.	Species	Chloroplast genome size (bp)	Non-coding size (bp)	Coding size (bp)	Clade
KJ614418	<i>Aegilops bicornis</i>	136,861	77,543	59,318	Clade I
KJ614416	<i>Aegilops longissima</i>	136,875	77,557	59,318	Clade I
KY636033	<i>Aegilops markgrafii</i>	136,063	76,745	59,318	Clade I
KJ614413	<i>Aegilops searsii</i>	136,870	77,552	59,318	Clade I
KJ614419	<i>Aegilops sharonensis</i>	136,867	77,549	59,318	Clade I
KJ614406	<i>Aegilops speltoides</i>	135,652	76,334	59,318	Clade I
KJ614405	<i>Aegilops speltoides</i> ssp. <i>ligustica</i>	135,660	76,343	59,317	Clade I
KJ614412	<i>Aegilops tauschii</i>	135,568	76,250	59,318	Clade I
KY636059	<i>Aegilops umbellulata</i>	136,028	76,710	59,318	Clade I
KY636056	<i>Aegilops umbellulata</i> ssp. <i>transcaucasica</i>	136,031	76,713	59,318	Clade I
KY636075	<i>Amblyopyrum muticum</i>	135,787	77,485	58,302	Clade I
KC912691	<i>Secale cereale</i>	135,564	77,109	58,455	Clade I
MH285856	<i>Taeniatherum caput-medusae</i>	136,861	77,537	59,324	Clade I
LC005977	<i>Triticum monococcum</i>	136,886	77,568	59,318	Clade I
KC912692	<i>Triticum monococcum</i> ssp. <i>aegilopoides</i>	136,870	75,821	61,049	Clade I
KJ614411	<i>Triticum urartu</i>	136,865	77,547	59,318	Clade I
MH285849	<i>Crithopsis delileana</i>	136,436	77,109	59,327	Clade I
MH285854	<i>Heterantherium piliferum</i>	136,768	77,490	59,278	Clade I
Average		136,362	77,053	59,308	
KX822019	<i>Pseudoroegneria libanotica</i>	135,026	75,711	59,315	Clade II
MH285855	<i>Pseudoroegneria spicata</i>	135,165	75,854	59,311	Clade II
MH285850	<i>Dasyphyrum villosum</i>	135,249	75,877	59,372	Clade II
MH331643	<i>Lophopyrum elongatum</i>	135,020	75,707	59,313	Clade II
MH331639	<i>Thinopyrum bessarabicum</i>	135,003	75,694	59,309	Clade II
Average		135,093	75,769	59,324	
KY126307	<i>Agropyron cristatum</i>	135,554	76,251	59,303	Clade III
MH285848	<i>Agropyron mongolicum</i>	135,547	76,244	59,303	Clade III
MH331642	<i>Australopyrum retrofractum</i>	135,417	76,114	59,303	Clade III
MH285851	<i>Eremopyrum distans</i>	135,589	76,248	59,341	Clade III
MH285852	<i>Eremopyrum tririceum</i>	135,554	76,283	59,271	Clade III
MH285853	<i>Henradia persica</i>	135,659	76,353	59,306	Clade III
Average		135,553	76,249	59,305	
MH331641	<i>Hordeum bogdanii</i>	136,968	77,672	59,296	Clade IV
KM974741	<i>Hordeum jubatum</i>	136,826	77,466	59,360	Clade IV
EF115541	<i>Hordeum vulgare</i>	136,462	77,153	59,309	Clade IV
KC912689	<i>Hordeum vulgare</i> ssp. <i>spontaneum</i>	136,043	76,945	59,098	Clade IV
Average		136,575	77,309	59,266	

Thinopyrum) and III (*Agropyron*, *Australopyrum*, *Eremopyrum*, and *Henradia*).

The amounts of IGSs showed extensive fluctuations and were attributed to major plastome size variation among the 34 taxa studied (Figures 3, 6), such as the 529-bp deletion between the *petN* gene and the *trnC* gene (Supplementary Figure 5) in the species within the clade II (*Pseudoroegneria*, *Dasyphyrum*, *Lophopyrum*, and *Thinopyrum*), and the 438-bp deletion between *psbE* and *petL* in the species within clade III (*Agropyron*, *Australopyrum*, *Eremopyrum*, and *Henradia*) (Supplementary Figure 6). These deletions might occur in the ancestral plastid genomes of species in clades II and III, inherited by and maintained in their offspring species. Variations in intergenic regions do not affect protein function and structure but cause variations in chloroplast genomes' size. Intergenic regions in

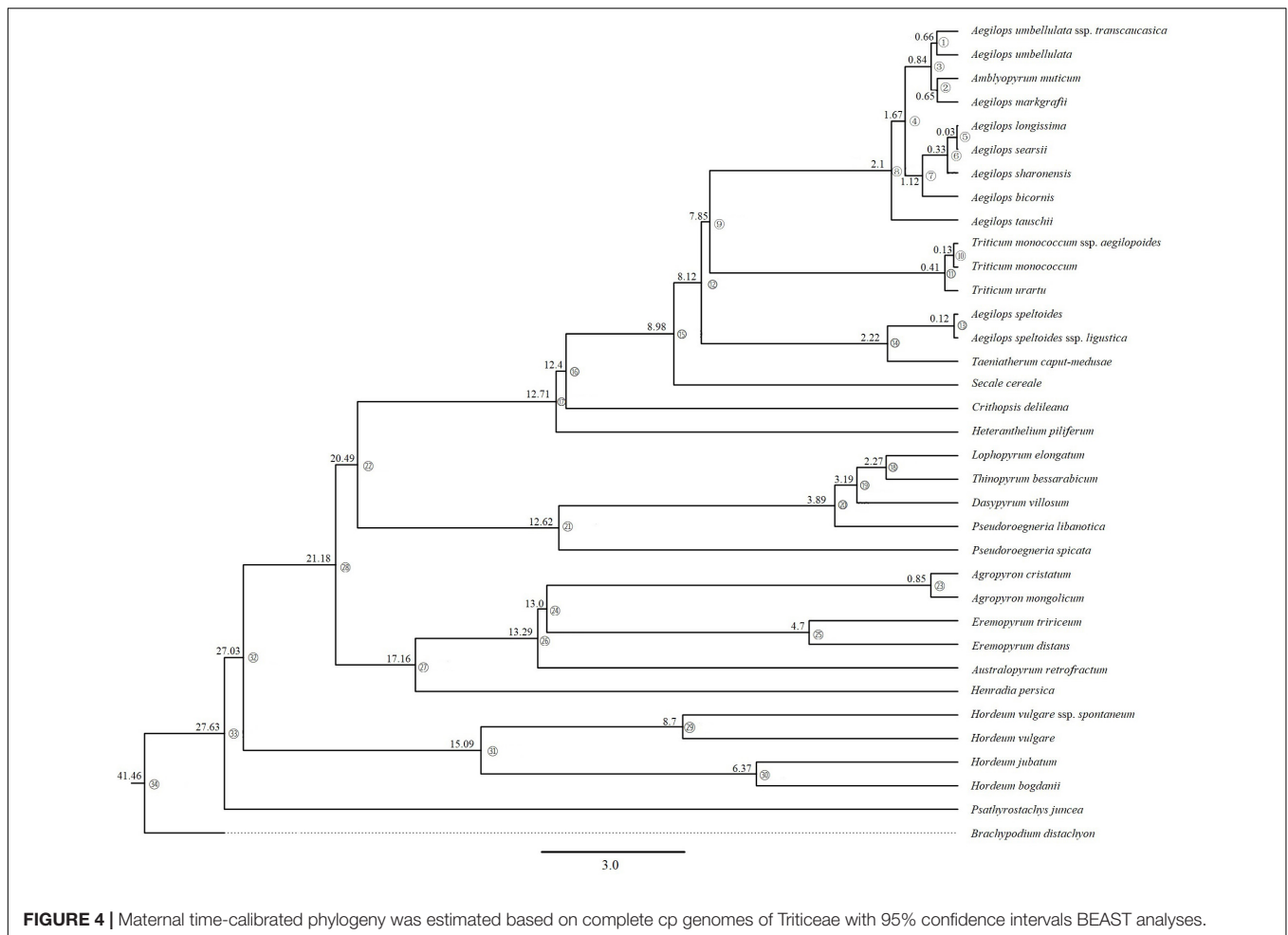
the cp genome have been utilized for systematic studies in diverse plants because they provide a wealth of information for distinguishing different species, such as *trnL-trnF* and *trnH-psbA* (Sha et al., 2014; Sevindik et al., 2021). Therefore, these intergenic regions (like *petN-trnC* and *psbE-petL*) can be used as an important tool to identify the maternal relationship among Triticeae species in the future.

The amount of introns, such as the intron of *atpF* (Supplementary Figure 16) and the intron of *rpl16* (Supplementary Figure 17), also showed extensive changes, in LSC, but no obvious patterns can be discerned in a phylogenetic context. The size variations in those introns were most likely the consequence of non-adaptive processes. However, random genetic drift might also play a central role in the shaping of plastome architecture (Smith, 2016).



Inverted repeat expansion and contraction were mostly associated with cp genome size variation (McCoy et al., 2008; Wu et al., 2009). Expansion and contraction in IRs of chloroplast genomes have been widely reported in various kinds of plants (Krause, 2011; Weng et al., 2014; Schwarz et al., 2015; Shrestha et al., 2019; He et al., 2020; Guo et al., 2021). We also found that cp genomes in the 34 Triticeae species exhibited obvious different IR sizes. A ~800-bp length variation in IR was detected between *trnI* and *trnL*, which mainly contained pseudogenes *ycf2* and *ycf15*. Wu et al. (2017) suggested that this deletion in the cp of *P. libanotica* was specific in Triticeae species. We confirmed that the loss of pseudogenes *ycf2* and *ycf15* occurred in the cp of *Pseudoroegneria*, *Dasyphyrum*, *Lophopyrum*, and *Thinopyrum*. The *Lophopyrum*, *Thinopyrum*, and *Dasyphyrum* genomes contributed a cytoplasm genome to the *Pseudoroegneria* species as a result of incomplete lineage sorting and/or chloroplast capture (Chen et al., 2020). A maternal donor of the species in clade II might have lost pseudogenes *ycf2*, *ycf15*, and adjacent gene spacers, and genetically transmitted to their offspring, leading to the IR size variation that ranged from 20,806 (*D. villosum*) to 20,866 bp (*P. spicata*), which was smaller than that in the other Triticeae species (>21 kb) (Figure 2). In some plants, *ycf2*-encoded protein and five related nuclear-encoded *FtsH* comprised a 2-MD complex, which can promote ATP synthesis (Kikuchi et al., 2018). The disappearance of *ycf2* might indicate that this function region has been transferred to a nuclear genome. Wu et al. (2017) found that this deletion

was exactly similar (identities = 99%) to a genomic scaffold of chromosome 3B of *T. aestivum*, and that it possesses high similarity (identify = 98%) with cp sequences in Pooideae species, but the specific function is still unknown. Although many species in different lineages contain an intact *ycf15* gene (encoding the beta subunit of acetyl-CoA carboxylase complex) and have been annotated in several sequenced chloroplast genomes, e.g., several *Asterids*, *Magnolia* (Kuang et al., 2011), and *Piper* (Cai et al., 2006), it is almost impossible to determine whether this gene is able to encode a functional protein or how it has evolved in angiosperms so far. The gene was first identified in the *Nicotiana* chloroplast genome (Shinozaki et al., 1986), and its similar expression was also reported in Solanaceae chloroplasts (Legen et al., 2002). However, the validity of *ycf15* as a protein-coding gene has long been questioned (Steane, 2005; Chumley et al., 2006). Its function was disabled in some basal angiosperms such, as *Amborella* (Goremykin et al., 2003) and *Nuphar* (Raubeson et al., 2007), monocots, most rosids, and some species in other lineages. It has been wholly lost in some other lineages, such as *Illicium*, *Acorus*, *Ceratophyllum*, and *Ranunculus* during their evolution processes. Transcriptome analyses revealed that *ycf15* has transcribed as a precursor polycistronic transcript that contained *ycf2*, *ycf15*, and antisense *trnL*-CAA. Pseudogene *ycf15* was mapped by multiple transcripts, which suggested that plastid DNA posttranscriptional splicing might involve a complex cleavage of non-functional genes (Shi et al., 2013).



A similar pseudogenetic elimination also occurred in the LSC of clade III species (*Agropyron*, *Eremopyrum*, *Australopyrum*, and *Henradia*), such as the *rpl23* pseudogene copy and the *accD* pseudogene between the *rbcl* gene and the *psaI* gene (Supplementary Figure 2). In the grass plastome, the *rpl23* gene is originally located in the IR region and encodes the functional ribosomal protein L23. Present studies have indicated that *rpl23* has been non-reciprocally translocated to a region downstream from *rbcl* (Bowman et al., 1988; Ogihara et al., 1988; He et al., 2017; Lencina et al., 2019). A non-reciprocal translocation of the *rpl23* gene occurred during the differentiation of *Poaceae* from its unknown ancestor (Katayama and Ogihara, 1996). In this study, we detected only ~260 bp of *rpl23* at the 3' end (complete gene size was 282 bp), which was inserted into the region between *rbcl* and *psaI* within the LSC of Triticeae species (except for *Aegilops tauschii*, *Agropyron*, *Australopyrum*, *Eremopyrum*, *Henradia*, and *S. cereale*). Meanwhile, sequence analysis detected another inserted pseudogene *accD* in this region, which is situated downstream of *rpl23*. In angiosperms, the *accD* gene encodes the protein of the acetyl-CoA carboxylase subunit D in the plastome. It is notable that the loss of *accD* is associated with hotspots of rearrangements in each of the families, introducing sensitivity to the herbicides quizalofop and

sethoxydim (Konishi and Sasaki, 1994), and causing an alteration of lipid metabolism in the plants. The loss of the *accD* gene had been found in four lineages of angiosperms, such as grasses (Hiratsuka et al., 1989; Maier et al., 1995; Katayama and Ogihara, 1996), *Campanulaceae* (Cosner et al., 1997), *Geraniaceae* (Palmer et al., 1987; Chumley et al., 2006), and *Aroideae* (Henriquez et al., 2021). Katayama and Ogihara (1996) considered *accD* loss prior to the divergence of the Poales. However, previous studies have shown the absence of the *accD* pseudogene in at least one species of Triticeae (Ogihara et al., 2002), and the presence of the *accD* pseudogene of up to 349 bp in *Secale* of the Triticeae species (Aagesen et al., 2005). The highly varied pattern of *accD* pseudogene presence or gene absence in members of the restiid and graminid clades might be due to the pseudogene being carried as the ancestral state throughout most of the divergence of the Poales (Harris et al., 2013). The mechanism for the insertion of variable sizes in this region is uncertain. In this study, the absence of the *accD* gene was only detected in species of clade III, *A. speltoides*, and *A. tauschii* in Triticeae (Supplementary Figure 3). The *accD* gene loss might accelerate gene relocations by unknown mechanisms, and various movements of genes might induce the sequential loss of *accD* (Lee et al., 2007).

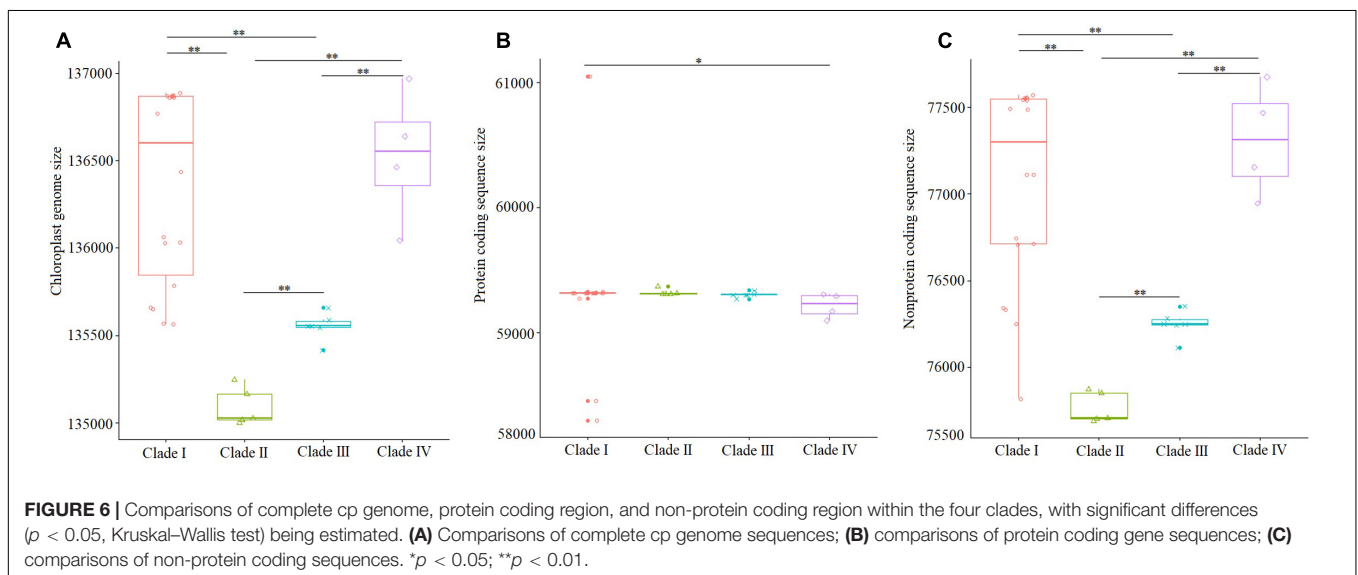
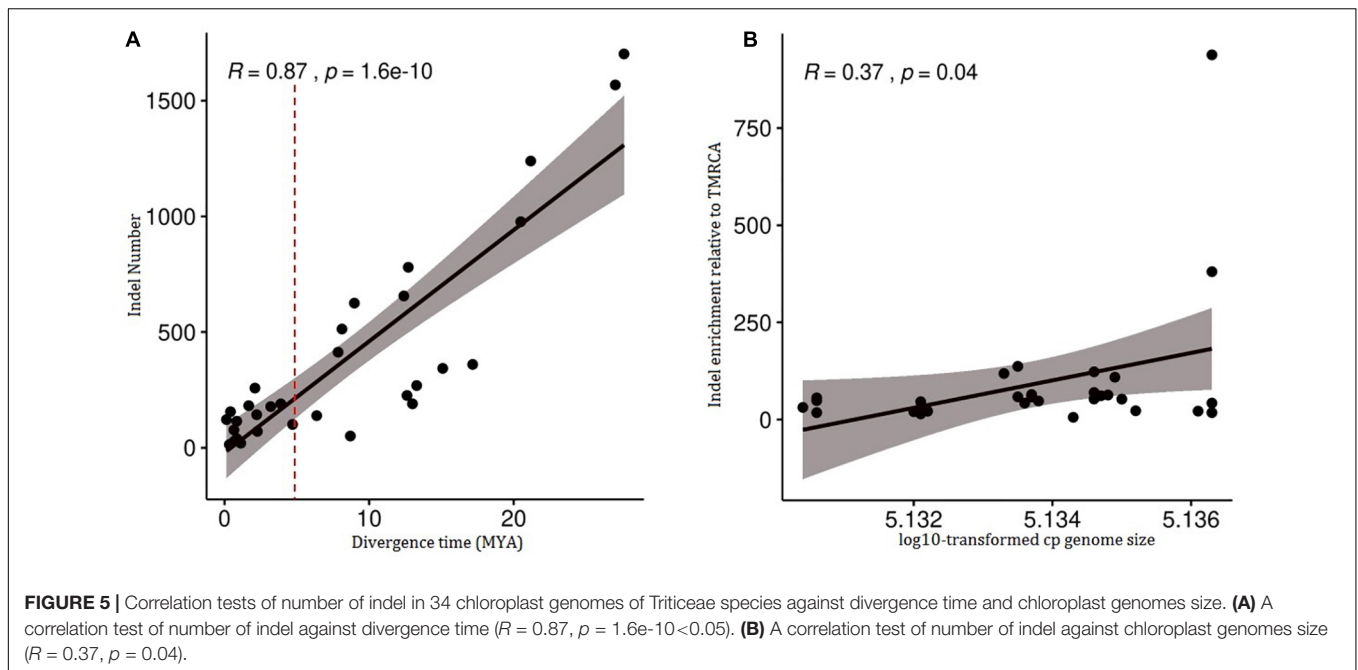
TABLE 3 | Divergence time and indels in every node of 34 Triticeae species in the phylogenomic tree.

Node	Divergence time (MYA)	Indel	CP genome size
1	0.66	28	136,030
2	0.65	77	135,925
3	0.84	115	135,977
4	1.67	182	136,423
5	0.03	–	136,873
6	0.33	14	136,871
7	1.12	20	136,868
8	2.1	258	136,328
9	7.85	413	136,464
10	0.13	122	136,878
11	0.41	156	136,874
12	8.12	513	136,383
13	0.12	–	135,656
14	2.22	143	136,058
15	8.98	625	136,332
16	12.4	656	136,338
17	12.71	780	136,362
18	2.27	71	135,012
19	3.19	178	135,091
20	3.89	190	135,075
21	12.62	226	135,093
22	20.49	977	136,086
23	0.85	39	135,551
24	13	190	135,561
25	4.7	101	135,572
26	13.29	269	135,532
27	17.16	360	135,553
28	21.18	1,239	135,976
29	8.7	51	136,253
30	6.37	139	136,804
31	15.09	343	136,528
32	27.03	1,568	136,043
33	27.63	1,702	136,059

The monogenomic Triticeae was diverged about 12–15 MYA to generate lineages, resulting in four branches of plastome species. Our ancestral genome reconstruction suggests that ancestral plastome species of clades II and III lost redundant sequences during divergence from Triticeae. The cp DNA sizes of species in the clades II and III (*Pseudoroegneria*, *Dasypyrum*, *Eremopyrum*, *Lophopyrum*, *Thinopyrum*, *Agropyron*, *Australopyrum*, and *Henradia*) evolved toward size reduction (Figures 3, 6) because of elimination of pseudogene and loss of long fragments (>200 bp) in IGS. Compared with species in clades I and IV (*Aegilops/Triticum* complex, *Taeniatherum*, *Secale*, *Crithopsis*, *Herterantherium*, and *Hordeum*) had larger cp genome size, and retained the invalid genes and lots of redundant fragments of IGS (Figures 3, 6). The cpDNA size reduction in the species of clades II and III (*Pseudoroegneria*, *Dasypyrum*, *Eremopyrum*, *Lophopyrum*, *Thinopyrum*, *Agropyron*, *Australopyrum*, and *Henradia*) might increase the efficiency of

genome replication. Genome reduction is speculated to be the result of a low-cost strategy that could facilitate rapid genome replication under disadvantageous environmental conditions (McCoy et al., 2008; Wu et al., 2009). The retained pseudogenes and replicates might participate in some nucleo-plasmic interactions to promote gene function. Although the *accD*, *ycf2*, and *ycf15* genes had been lost from the plastid genome several times in angiosperms, their functions were fulfilled by nuclear copies (Nakkaew et al., 2008; Huang et al., 2017; Wu et al., 2017).

Environmental conditions were thought to influence organelle DNA architecture. For example, plastid genomic compaction in the endolithic ulvophyte seaweed (*Ostreobium quekettii*) and the palmophylalean green alga (*Verdigellas peltata*) was caused primarily by adaptation to low light conditions (Marcelino et al., 2016). The large size variation in major lineages and their subclades is most likely the consequence of adaptive processes since those variations are highly positively correlated with divergence time (Figure 5). Plastid genomes smaller than 120,000 bp were detected mostly in non-photosynthetic angiosperms that had a deletion of several genes during their diversification (Wicke et al., 2011, 2016). The diversification of high plant species has been found to be strongly linked to climate fluctuations (Jaramillo et al., 2006; Hoorn et al., 2010). During the late Miocene (5–10 MYA), the atmospheric CO₂ level was decreased to the bottom after the mid-Miocene climate optimum (14–16 MYA) (Tripathi et al., 2009), resulting in climate change from greenhouse to the icehouse. In order to adapt to an extremely cold and low CO₂ concentration climate, most species will reduce metabolism to maintain survival, and correspondingly, the expression of genes related to photosynthesis would be reduced, since CO₂ is required for photosynthesis. The species in the clade I from genera *Aegilops*, *Triticum*, *Secale*, *Taeniatherum*, *Crithopsis*, and *Heterantherium* were restrictedly distributed in the Mediterranean and adjacent regions (Sakamoto, 1973; Hsiao et al., 1995), where there are hot, dry environments such as the deserts of Mediterranean regions. The sizes of complete cp genome and non-protein coding sequences in Mediterranean lineage were larger than those of other genera as the ancestral species evolved and diverged 0.03–12.71 MYA (Figure 3), and variations appeared most frequently about 5 MYA (Figure 5). The main diversification of Mediterranean lineage in Triticeae occurred about 9 MYA when Mediterranean climates are thought to have arisen. In the late Miocene, when atmospheric CO₂ concentrations and temperatures were extremely low, the Mediterranean climate undoubtedly provided a good shelter for some plants. The development of the Mediterranean climate can be seen as the opening of a new and novel climatic niche, to which lineages have adapted and speciated, by accumulating morphological change in other climate zones (Yesson and Culham, 2006). It is, thus, likely that climate oscillations during the late Miocene, especially the establishment of the Mediterranean climate, might have promoted the Mediterranean lineage of Triticeae rapid diversification and adaptation, and have continued to diversify from the Quaternary to the present (Fan et al., 2013), which



resulted in plastid genome and nuclear genome change to adapt to climate oscillations. We suggested that the dynamically reduced/enlarged cpDNAs of Triticeae might result from the adaptation to historical climate changes. However, we still need more molecular evidence to determine the role of natural selection in chloroplast genome evolution during the diversification of Triticeae.

CONCLUSION

In this study, we detected gene loss/pseudonization, indels, and intron variation, variation in cp genome sequence size, and

expansion/contraction in IRs among 34 chloroplast genomes of monogenomic Triticeae species. We found that the cp genome sequence size variation was mainly caused by the size of non-protein coding sequences. The monogenomic Triticeae diverged about 12–15 MYA, which resulted in four stem branches of plastome species. Losses of a series of invalid genes or sequence fragments have no effect on genomic function in these plants, which might be an evolutionary mechanism to increase the efficiency of genome replication. According to the distribution and habitat of the species, the species in the Mediterranean region (in clade I, *Aegilops/Triticum* complex, *Taeniatherum*, *Secale*, *Crithopsis*, and *Herteranthelium*) might have experienced the change in the Mediterranean climate and expanded the cp

genome significantly. Our results enhance the understanding of the complexity and evolution of Triticeae cp DNAs.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: the chloroplast genome sequences in this study, KJ614418, KJ614416, KY636033, KJ614413, KJ614419, KJ614406, KJ614405, KJ614412, KY636059, KY636056, KY126307, MH285848, KY636075, MH331642, MH285849, MH285850, MH285851, MH285852, MH285853, MH285854, MH331641, KM974741, EF115541, KC912689, MH331643, MH331640, KX822019, MH285855, KC912691, MH285856, MH331639, LC005977, KC912692, KJ614411, and EU325680 are available in NCBI.

AUTHOR CONTRIBUTIONS

NC and XF designed the experiments. NC carried out the experiments, analyzed the experimental results and data, and developed the analysis tools. L-NS, Y-LW, L-JY, YZ, YW, D-DW, H-YK, H-QZ, Y-HZ, and G-LS assisted in writing the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

- Aagesen, L., Petersen, G., and Seberg, O. (2005). Sequence length variation, indel costs, and congruence in sensitivity analysis. *Cladistics* 21, 15–30. doi: 10.1111/j.1096-0031.2005.00053.x
- Amiryousefi, A., Hyvonen, J., and Pocza, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34, 3030–3031. doi: 10.1093/bioinformatics/bty220
- Barkworth, M. E. (1998). *Grasses of the Tribe Hordeae in North America*. 3. *Comments. Botanical Electronic News* 199. Available online at: <https://www.ou.edu/cas/botany-micro/ben/ben199.html>.
- Bedoya, A. M., Ruhfel, B. R., Philbrick, C. T., Madriñán, S., Bove, C. P., Mesterházy, A., et al. (2019). Plastid genomes of five species of riverweeds (Podostemaceae): structural organization and comparative analysis in Malpighiales. *Front. Plant Sci.* 10:1035. doi: 10.3389/fpls.2019.01035
- Birky, C. W., Maruyama, T., and Fuerst, P. (1983). An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* 103, 513–527. doi: 10.1093/genetics/103.3.513
- Bortiri, E., Coleman-Derr, D., Lazo, G. R., Anderson, O. D., and Gu, Y. Q. (2008). The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Res. Notes* 1:61. doi: 10.1186/1756-0500-1-61
- Bowman, C. M., Barker, R. F., and Dyer, T. A. (1988). In wheat ctDNA, segments of ribosomal protein genes are dispersed repeats, probably conserved by nonreciprocal recombination. *Curr. Genet.* 14, 127–136. doi: 10.1007/BF00569336
- Brouard, J. S., Otis, C., Lemieux, C., and Turmel, M. (2010). The exceptionally large chloroplast genome of the green alga *Floydiella terrestris* illuminates the evolutionary history of the *Chlorophyceae*. *Genome Biol. Evol.* 2, 240–256. doi: 10.1093/gbe/evq014
- Cai, Z., Penafior, C., Kuehl, J. V., Leebens-Mack, J., Carlson, J. E., Claude, W. D., et al. (2006). Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. *BMC Evol. Biol.* 6:77. doi: 10.1186/1471-2148-6-77

FUNDING

This study was funded by the National Natural Science Foundation of China (Grant No. 31870360) and the Second Tibetan Plateau Science Expedition and Research Program (STEP) (No. 2019QZKK0303).

ACKNOWLEDGMENTS

We are very grateful to the American National Plant Germplasm System (Pullman, WA, United States) for providing the part seed materials.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.741063/full#supplementary-material>

Supplementary Figure 1 | Comparison of the chloroplast genome. Alignment of the cp genome sequences of 34 Triticeae species and *Brachypodium distachyon* generated with mVISTA. Gray arrows indicate the position and direction of each gene. Red and blue areas indicate intergenic and genic regions, respectively.

Supplementary Figures 2–17 | Indels in protein coding genes, gene loss/pseudonization, intron variation, and intergenic sequence (IGS) of 34 Triticeae diploids cp genomes.

- Chen, N., Chen, W. J., Yan, H., Wang, Y., Kang, H. Y., Zhang, H. Q., et al. (2020). Evolutionary patterns of plastome uncover diploid-polyploid maternal relationships in *Triticeae*. *Mol. Phylogenet. Evol.* 149:106838. doi: 10.1016/j.ympev.2020.106838
- Chen, N., Sha, L. N., Dong, Z. Z., Tang, C., Wang, Y., Kang, H. Y., et al. (2017). Complete structure and variation of the chloroplast genome of *Agropyron cristatum* (L.) Gaertn. *Gene* 640, 86–96. doi: 10.1016/j.gene.2017.10.009
- Choi, K. S., and Park, S. J. (2015). The complete chloroplast genome sequence of *Aster spathulifolius*, (*Asteraceae*); genomic features and relationship with *Asteraceae*. *Gene* 572, 214–221. doi: 10.1016/j.gene.2015.07.020
- Chumley, T. W., Palmer, J. D., Mower, J. P., Fourcade, H. M., Calie, P. J., Boore, J. L., et al. (2006). The complete chloroplast genome sequence of *Pelargonium × hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol.* 23, 2175–2190. doi: 10.1093/molbev/msl089
- Cosner, M. E., Jansen, R. K., Palmer, J. D., and Downie, S. R. (1997). The highly rearranged chloroplast genome of *Trachelium caeruleum* Cam-panulaceae: multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Curr Genet.* 31, 419–429. doi: 10.1007/s002940050225
- Dewey, D. R. (1984). “The genomic system of classification as a guide to intergeneric hybridization with the perennial *Triticeae*,” in *Gene Manipulation in Plant Improvement*, ed. J. P. Gustafson (New York, NY: Columbia University Press), 209–279.
- Drummond, A. J., and Rambaut, A. (2007). BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:54. doi: 10.1186/s12879-015-0770-x
- Fan, X., Sha, L. N., Yu, S. B., Wu, D. D., Chen, X. H., Zhuo, X. F., et al. (2013). Phylogenetic reconstruction and diversification of the *Triticeae* (*Poaceae*) based on single-copy nuclear *Acc1* and *Pgk1* gene data. *Biochem. Syst. Ecol.* 50, 346–360. doi: 10.1016/j.bse.2013.05.010
- Finarelli, J. A., and Flynn, J. J. (2006). Ancestral state reconstruction of body size in the *Caniformia* (Carnivora, *Mammalia*): the effects of incorporating data

- from the fossil record. *Syst. Biol.* 55, 301–313. doi: 10.1080/10635150500541698
- Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32(Suppl._2), W273–W279. doi: 10.1093/nar/gkh458
- Goremykin, V. V., Hirsch-Ernst, K. I., Wölfl, S., and Hellwig, F. H. (2003). Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol. Biol. Evol.* 20, 1499–1505. doi: 10.1093/molbev/msg159
- Gornicki, P., Zhu, H., Wang, J., Challa, G. S., Zhang, Z., Gill, B. S., et al. (2014). The chloroplast view of the evolution of polyploid wheat. *New Phytol.* 204, 704–714. doi: 10.1111/nph.12931
- Guo, Y. Y., Yang, J. X., Bai, M. Z., Zhang, G. Q., and Liu, Z. J. (2021). The chloroplast genome evolution of Venus slipper (*Paphiopedilum*): IR expansion, SSC contraction, and highly rearranged SSC regions. *BMC Plant Biol.* 21:248. doi: 10.21203/rs.3.rs-257472/v1
- Harris, M. E., Meyer, G., Vandergon, T., and Vandergon, V. O. (2013). Loss of the acetyl-CoA carboxylase (*accD*) gene in Poales. *Plant Mol./ Biol. Rep.* 31, 21–31. doi: 10.1007/s11105-012-0461-3
- He, L., Qian, J., Li, X., Sun, Z., Xu, X., and Chen, S. (2017). Complete chloroplast genome of medicinal plant *Lonicera japonica*: genome rearrangement, intron gain and loss, and implications for phylogenetic studies. *Molecules* 22:249. doi: 10.3390/molecules22020249
- He, S., Yang, Y., Li, Z., Wang, X., Guo, Y., and Wu, H. (2020). Comparative analysis of four *Zantedeschia* chloroplast genomes: expansion and contraction of the IR region, phylogenetic analyses and SSR genetic diversity assessment. *PeerJ* 8:e9132. doi: 10.7717/peerj.9132
- Henriquez, C. L., Mehmood, F., Hayat, A., Sammad, A., Waseem, S., Waheed, M. T., et al. (2021). Chloroplast genome evolution in the *Dracunculaceae* clade (*Aroideae*, *Araceae*). *Genomics* 113, 183–192. doi: 10.1016/j.ygeno.2020.12.016
- Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., et al. (1989). The complete sequence of the rice *Oryza sativa* chloroplast genome—intermolecular recombination between distinct transfer-RNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* 217, 185–194. doi: 10.1007/BF02464880
- Hoorn, C., Wesselingh, F. P., ter Steege, H., Bermudez, M. A., Mora, A., Sevink, et al. (2010). Amazonia through time: andean uplift, climate change, landscape evolution, and biodiversity. *Science* 330, 927–931.
- Hoot, S. B., and Palmer, J. D. (1994). Structural rearrangements, including parallel inversions, within the chloroplast genome of *Anemone* and related genera. *J. Mol. Evol.* 38, 274–281. doi: 10.1007/BF00176089
- Hsiao, C., Chatterton, N. J., Asay, K. H., and Jensen, K. B. (1995). Phylogenetic relationships of the monogenomic species of the wheat tribe, *Triticeae* (*Poaceae*), inferred from nuclear rDNA (internal transcribed spacer) sequences. *Genome* 38, 211–223. doi: 10.1139/g95-026
- Huang, Y. Y., Cho, S. T., Haryono, M., and Kuo, C. H. (2017). Complete chloroplast genome sequence of common bermudagrass (*Cynodon dactylon* (L.) Pers.) and comparative analysis within the family *Poaceae*. *PLoS One* 12:e0179055. doi: 10.1371/journal.pone.0179055
- Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754
- Jansen, R. K., Wojciechowski, M. F., Sanniyasi, E., Lee, S. B., and Daniell, H. (2008). Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (*Leguminosae*). *Mol. Phylogenet. Evol.* 48, 1204–1217. doi: 10.1016/j.ympev.2008.06.013
- Jaramillo, C., Rueda, M. J., and Mora, G. (2006). Cenozoic plant diversity in the Neotropics. *Science* 311, 1893–1896. doi: 10.1126/science.1121380
- Katayama, H., and Ogihara, Y. (1996). Phylogenetic affinities of the grasses to other monocots as revealed by molecular analysis of chloroplast DNA. *Curr. Genet.* 29, 572–581. doi: 10.1007/BF02426962
- Katoh, K., and Toh, H. (2010). Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* 26, 1899–1900. doi: 10.1093/bioinformatics/btq224
- Kellogg, E. A., Appels, R., and Mason-Gamer, R. J. (1996). When genes tell different stories: the diploid genera of *Triticeae* (Gramineae). *Syst. Bot.* 21, 321–347. doi: 10.2307/2419662
- Kikuchi, S., Asakura, Y., Imai, M., Nakahira, Y., Kotani, Y., Hashiguchi, Y., et al. (2018). A Ycf2-FtsHi heteromeric AAA-ATPase complex is required for chloroplast protein import. *Plant Cell* 30, 2677–2703. doi: 10.1105/tpc.18.00357
- Konishi, T., and Sasaki, Y. (1994). Compartmentalization of two forms of acetyl CoA carboxylase in plants and the origin of their tolerance toward herbicides. *Proc. Natl. Acad. Sci. U.S.A.* 91, 3598–3601.
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455. doi: 10.1093/bioinformatics/btz305
- Krause, K. (2011). Piecing together the puzzle of parasitic plant plastome evolution. *Planta* 234, 647–656. doi: 10.1007/s00425-011-1494-9
- Kuang, D. Y., Wu, H., Wang, Y. L., Gao, L. M., Zhang, S. Z., and Lu, L. (2011). Complete chloroplast genome sequence of *Magnolia kwangsiensis* (*Magnoliaceae*): implication for DNA barcoding and population genetics. *Genome* 54, 663–673. doi: 10.1139/g11-026
- Kugita, M., Kaneko, A., Yamamoto, Y., Takeya, Y., Matsumoto, T., and Yoshinaga, K. (2003). The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. *Nucleic Acids Res.* 31, 716–721. doi: 10.1093/NAR/GKG155
- Lee, H. L., Jansen, R. K., Chumley, T. W., and Kim, K. J. (2007). Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (*Oleaceae*) are due to multiple, overlapping inversions. *Mol. Biol. Evol.* 24, 1161–1180. doi: 10.1093/molbev/msm036
- Legen, J., Kemp, S., Krause, K., Profanter, B., Herrmann, R. G., and Maier, R. M. (2002). Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries. *Plant J.* 31, 171–188. doi: 10.1046/j.1365-313X.2002.01349.x
- Lencina, F., Landau, A. M., Petterson, M. E., Pacheco, M. G., Kobayashi, K., and Prina, A. R. (2019). The *rpl23* gene and pseudogene are hotspots of illegitimate recombination in barley chloroplast mutator seedlings. *Sci. Rep.* 9, 1–13. doi: 10.1038/s41598-019-46321-6
- Löve, Á (1984). Conspectus of the *Triticeae*. *Feddes Rep.* 95, 425–521. doi: 10.1002/fedr.4910950702
- Maddison, W. P., and Maddison, D. R. (2021). *Mesquite: A Modular System for Evolutionary Analysis. Version 3.7*. Available online at: <http://mesquiteproject.org>.
- Maier, R. M., Neckermann, K., Igloi, G. L., and Kossel, H. (1995). Complete sequence of the maize chloroplast genome—gene content, hotspots of divergence and fine-tuning of genetic information by transcript editing. *J. Mol. Biol.* 251, 614–628. doi: 10.1006/jmbi.1995.0460
- Marcelino, V. R., Cremen, M. C. M., Jackson, C. J., Larkum, A. W., and Verbruggen, H. (2016). Evolutionary dynamics of chloroplast genomes in low light: a case study of the endolithic green alga *Ostreobium quekettii*. *Genome Biol. Evol.* 8, 2939–2951. doi: 10.1093/gbe/evw206
- Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K. S., et al. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345:1250092. doi: 10.1126/science.1250092
- Martin, T., Oswald, O., and Graham, I. A. (2002). *Arabidopsis* seedling growth, storage lipid mobilization and photosynthetic gene expression are regulated by carbon: nitrogen availability. *Plant Physiol.* 128, 472–481. doi: 10.1104/pp.010475
- Matsuoka, T., Omata, K., Kanda, H., and Tachi, K. (2002). “A study of wave energy conversion systems using ball screws— Comparison of output characteristics of the fixed type and the floating type,” in *Proceedings of the International Conference on Offshore and Polar Engineering*, Kitakyushu, 581–585.
- McCoy, S. R., Kuehl, J. V., Boore, J. L., and Raubeson, L. A. (2008). The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. *BMC Evol. Biol.* 8:130. doi: 10.1186/1471-2148-8-130
- Middleton, C. P., Senerchia, N., Stein, N., Akhunov, E. D., Keller, B., Wicker, T., et al. (2014). Sequencing of chloroplast genomes from wheat, barley, rye and

- their relatives provides a detailed insight into the evolution of the *Triticeae* tribe. *PLoS One* 9:e0085761. doi: 10.1371/journal.pone.0085761
- Nakkaew, A., Chotigeat, W., Eksomtramage, T., and Phongdara, A. (2008). Cloning and expression of a plastid-encoded subunit, betacarboxyltransferase gene (*accD*) and a nuclear-encoded subunit biotin carboxylase of acetyl-CoA carboxylase from oil palm (*Elaeis guineensis* Jacq.). *Plant Sci.* 175, 497–504. doi: 10.1016/j.plantsci.2008.05.023
- Nock, C. J., Waters, D. L., Edwards, M. A., Bowen, S. G., Rice, N., Cordeiro, G. M., et al. (2011). Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol. J.* 9, 328–333. doi: 10.1111/j.1467-7652.2010.00558.x
- Ogihara, Y., Isono, K., Kojima, T., Endo, A., Hanaoka, M., Shiina, T., et al. (2002). Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. *Mol. GenGenom* 266, 740–746. doi: 10.1007/s00438-001-0606-9
- Ogihara, Y., Terachi, T., and Sasakuma, T. (1988). Intramolecular recombination of chloroplast genome mediated by a short direct-repeat sequence in wheat species. *Proc. Natl. Acad. Sci. U.S.A.* 85, 8573–8577. doi: 10.1073/pnas.85.22.8573
- Palmer, J. D. (1985). “Evolution of chloroplast and mitochondrial DNA in plants and algae,” in *Molecular Evolutionary Genetics*, ed. R. MacIntyre (New York, NY: Plenum Press), 131–240. doi: 10.1007/978-1-4684-4988-4_3
- Palmer, J. D., Nugent, J. M., and Herbon, L. A. (1987). Unusual structure of geranium chloroplast DNA—a triple-sized inverted repeat, extensive gene duplications, multiple inversions, and two repeat families. *Proc. Natl. Acad. Sci. U.S.A.* 84, 769–773. doi: 10.2307/28809
- Parks, M., Cronn, R., and Liston, A. (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* 7:84. doi: 10.1186/1741-7007-7-84
- Rambaut, A., and Drummond, A. J. (2018). *Tracer v1.4: MCMC Trace Analyses Tool*. Available online at: <http://beast.bio.ed.ac.uk/Tracer>.
- Raubeson, L. A., Peery, R., Chumley, T. W., Dziubek, C., Fourcade, H. M., Boore, J. L., et al. (2007). Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genomics* 8:174. doi: 10.1186/1471-2164-8-174
- Raubeson, L. A., and Jansen, R. K. (2005). “Chloroplast genomes of plants,” in *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*, Vol. 3, ed. R. J. Henry (Cambridge, MA: CABI), 45–68. doi: 10.1079/9780851999043.0045
- Saarela, J. M., Wysocki, W. P., Barrett, C. F., Soreng, R. J., Davis, J. I., Clark, L. G., et al. (2015). Plastid phylogenomics of the cool-season grass subfamily: clarification of relationships among early-diverging tribes. *AoB Plants* 7:lv046. doi: 10.1093/aobpla/plv046
- Sakamoto, S. (1973). Patterns of phylogenetic differentiation in the tribe *Triticeae*. *Seiken Zihō* 24, 11–31.
- Saski, C., Lee, S. B., Fjellheim, S., Guda, C., Jansen, R. K., Luo, H., et al. (2007). Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theoret. Appl. Genet.* 115, 571–590. doi: 10.1007/s00122-007-0567-4
- Schwarz, E. N., Ruhlman, T. A., Sabir, J. S., Hajrah, N. H., Alharbi, N. S., Al-Malki, A. L., et al. (2015). Plastid genome sequences of legumes reveal parallel inversions and multiple losses of *rps16* in papilionoids. *J. Syst. Evol.* 53, 458–468. doi: 10.1111/jse.12179
- Sevindik, E., Aydoan, M., and Yavuz, M. (2021). Phylogenetic analysis of the genus *Conringia* Heist. Ex Fabr. (Brassicaceae) in Turkey based on nuclear (nrITS) and chloroplast (trnL-F) DNA sequences. *Notulae Sci. Biol.* 13:11034. doi: 10.15835/nsb13311034
- Sha, L., Fan, X., Zhang, H., Kang, H., Wang, Y., Wang, X., et al. (2014). Phylogenetic relationships in *Leymus* (*Triticeae*; *Poaceae*): evidence from chloroplast *trnH-psbA* and mitochondria *coxII* intron sequences. *J. Sys. Evol.* 52, 722–734. doi: 10.1111/jse.12097
- Shi, C., Liu, Y., Huang, H., Xia, E. H., Zhang, H. B., and Gao, L. Z. (2013). Contradiction between plastid gene transcription and function due to complex posttranscriptional splicing: an exemplary study of *ycf15* function and evolution in angiosperms. *PLoS One* 8:e0059620. doi: 10.1371/journal.pone.0059620
- Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., et al. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* 5, 2043–2049. doi: 10.1002/j.1460-2075.1986.tb04464.x
- Shrestha, B., Weng, M. L., Theriot, E. C., Gilbert, L. E., Ruhlman, T. A., Krosnick, S. E., et al. (2019). Highly accelerated rates of genomic rearrangements and nucleotide substitutions in plastid genomes of *Passiflora* subgenus *Decaloba*. *Mol. Phylogenet. Evol.* 138, 53–64. doi: 10.1016/j.ympev.2019.05.030
- Smith, D. R. (2016). The mutational hazard hypothesis of organelle genome evolution: 10 years on. *Mol. Ecol.* 25, 3769–3775. doi: 10.1111/mec.13742
- Smith, D. R., Hamaji, T., Olson, B. J., Durand, P. M., Ferris, P., Michod, R. E., et al. (2013). Organelle genome complexity scales positively with organism size in volvocine green algae. *Mol. Biol. Evol.* 30, 793–797. doi: 10.1093/molbev/mst002
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stearns, D. A. (2005). Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). *DNA Res.* 12, 215–220. doi: 10.1093/dnares/dsi006
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Tripati, A. K., Roberts, C. D., and Eagle, R. A. (2009). Coupling of CO₂ and ice sheet stability over major climate transitions of the last 20 million years. *Science* 326, 1394–1397.
- Turmel, M., Otis, C., and Lemieux, C. (2015). Dynamic evolution of the chloroplast genome in the green algal classes Pedinophyceae and Trebouxiophyceae. *Genome Biol. Evol.* 7, 2062–2082. doi: 10.1093/gbe/evv130
- Turmel, M., Otis, C., and Lemieux, C. (2017). Divergent copies of the large inverted repeat in the chloroplast genomes of ulvophycean green algae. *Sci. Rep.* 7, 1–14. doi: 10.1038/s41598-017-01144-1
- Walker, J. F., Zanis, M. J., and Emery, N. C. (2014). Comparative analysis of complete chloroplast genome sequence and inversion variation in *Lasthenia burkei* (*Madieae*, *Asteraceae*). *Am. J. Bot.* 101, 722–729. doi: 10.3732/ajb.1400049
- Wang, R. R.-C., Bothmer, R. V., Dvorak, J., Fedak, G., Linde-Laursen, I., and Muramatsu, M. (1994). “Genome symbols in the *Triticeae* (*Poaceae*),” in *Proceedings of the 2nd International Triticeae Symposium*, eds R. R.-C. Wang, K. B. Jensen, and C. Jaussi (Logan), 29–34.
- Weng, M. L., Blazier, J. C., Govindu, M., and Jansen, R. K. (2014). Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol. Biol. Evol.* 31, 645–659. doi: 10.1093/molbev/mst257
- West, J. G., McIntyre, C. L., and Appels, R. (1988). Evolution and systematic relationships in the *Triticeae* (*Poaceae*). *Plant Syst. Evol.* 160, 1–28. doi: 10.1007/BF00936706
- Wicke, S., Müller, K. F., DePamphilis, C. W., Quandt, D., Bellot, S., and Schneeweiss, G. M. (2016). Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9045–9050. doi: 10.1073/pnas.1607561113
- Wicke, S., Schneeweiss, G. M., Depamphilis, C. W., Müller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4
- Wu, C. S., Lai, Y. T., Lin, C. P., Wang, Y. N., and Chaw, S. M. (2009). Evolution of reduced and compact chloroplast genomes (cpDNAs) in angiosperms: selection towards a lower cost strategy. *Mol. Phylogenet. Evol.* 52, 115–124. doi: 10.1016/j.ympev.2008.12.026
- Wu, C. S., Lin, C. P., Hsu, C. Y., Wang, R. J., and Chaw, S. M. (2011). Comparative chloroplast genomes of pinaceae: insights into the mechanism of diversified genomic organizations. *Genome Biol. Evol.* 3, 309–319. doi: 10.1093/gbe/evr026
- Wu, D. D., Sha, L. N., Tang, C., Fan, X., Wang, Y., Kang, H. Y., et al. (2017). The complete chloroplast genome sequence of *Pseudoroegneria libanotica*, genomic

- features, and phylogenetic relationship with *Triticeae* species. *Biol. Plant.* 62, 231–240. doi: 10.1007/s10535-017-0759-y
- Yesson, C., and Culham, A. (2006). Phyloclimatic modeling: combining phylogenetics and bioclimatic modeling. *Syst. Biol.* 55, 785–802. doi: 10.1080/1063515060081570
- Zheng, X. M., Wang, J., Feng, L., Liu, S., Pang, H., Qi, L., et al. (2017). Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants. *Sci. Rep.* 7:1555. doi: 10.1038/s41598-017-01518-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Chen, Sha, Wang, Yin, Zhang, Wang, Wu, Kang, Zhang, Zhou, Sun and Fan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.