INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
20–1806/2021/25–6–1203–1212
DOI: 10.17957/IJAB/15.1700
http://www.fspublishers.org

Full Length Article

Comparative and Phylogenetic Analysis of Complete Chloroplast Genome Sequences of Two Wildrye Grasses *Elymus sibiricus* and *E. nutans* (Triticeae, Poaceae)

Qingqing Yu[1†], Zhechuan Liu[1†], Yi Xiong[1], Yanli Xiong[1], Cong Nie[1], Haidong Gao[3], Wenhui Liu[2] and Xiao Ma[1*]
[1]*College of Animal Science and Technology, Sichuan Agricultural University, Chengdu 611130, P. R. China*
[2]*Qinghai Academy of Animal Science and Veterinary Medicine, Key Laboratory of Superior Forage Germplasm in the Qinghai-Tibetan Plateau, Xi-ning, P. R. China*
[3]*Genepioneer Biotechnologies Co. Ltd, Nanjing, 210023, P. R. China*
[*]For correspondence: maroar@126.com
[†]Contributed equally to this work and are co-first authors
Received 30 December 2020; Accepted 01 March 2021; Published 10 May 2021

Abstract

Due to outstanding characteristics such as stress resistance and high biomass production, *Elymus sibiricus* (StH genomes) and *E. nutans* (StHY genomes) are regarded as ecologically important perennial bunchgrass species belonging to *Elymus* genus of tribe Triticeae (Poaceae), which were widely used to promote the restoration of degraded grassland in the eastern Tibetan Plateau. In this study, the complete chloroplast (cp) genome of *E. sibiricus* and *E. nutans* were sequenced and annotated with *de novo* analysis, to clarify their inter-species variation and their evolutionary relationships with relative species. The result showed that both two whole cp genomes shared a typical quadripartite structure, the cp genome length of *E. sibiricus* and *E. nutans* were 135,075 bp and 135,060 bp, respectively. Three genes *tRNA-CGA*, *tRNA-CGU*, and *tRNA-CGU* were unique in *E. sibiricus* while the gene *ycf1* (hypothetical chloroplast reading frame no. 1) was only found in *E. nutans*. The identification of hotspot regions (*tRNA-GUC~psbM*, *tRNA-UAA~ndhJ*, *rbcL~psaI*, *rpl33~rps18*) between the two cp genomes would be pertinent to the development of barcode marker of these two *Elymus* species. Comparative cp genome analysis and phylogenetic relationships further confirmed that *Pseudoroegneria* were putative matrilineal donors of St genome of *Elymus* species at plastome level. Whole cp genomes could be used as an effective barcode for species identification or for developing specific markers, which is essential useful for the evolutionary history and conservation of *Elymus* species. © 2021 Friends Science Publishers

Keywords: *Elymus* spp.; Chloroplast genome; Hotspot regions; Phylogenetic analysis

Introduction

Elymus L. is the largest genus of approximate 150 species of perennial grasses in the Triticeae tribe (Poaceae), also called wildrye and their species are widely distributed in most of the temperate regions in the world (Zhang *et al*. 2019). Given the high biomass, good forage quality and excellent tolerance to multiple biotic and abiotic stresses, the *Elymus* species are of great importance to the artificial grassland construction and degraded grassland restoration in northwestern China (Qiao *et al*. 2006). Furthermore, the excellent stress-resistance genes derived from *Elymus* species could be transferred to the related cereal crops for genetic improvement. *E. sibiricus*, with the genome constitution of StStHH (2n = 4x = 28), along with *E. nutans* (StStHHYY, 2n=6x=42), are the two most common perennial grasses species and widely used in forage production and restoring degraded grassland in the eastern Qinghai–Tibet Plateau (Zhang *et al.* 2016b). *E. sibiricus* usually has a lower drought resistance and higher biomass yield than *E. nutans.* However, the high morphological similarity and niche overlaps limit their germplasm identification and further hinders seed production and promotion (Lei *et al.* 2014). The sequencing of plastid genome via the next-generation sequencing technology (NGS) could provide a convenient and cost-effective approach to develop molecular markers, which was a potential tool for germplasms/species identification, evolutionary study, genetic relationship evaluation, and haplotype division derived from plastome (Li *et al.* 2014).

Chloroplast (cp) is an important component of plant organelles and photosynthetic organs (Hong *et al.* 2020). The cp genome was reported to be consisted of a typical quadripartite structure with a large single-copy region (LSC),

To cite this paper: Yu Q, Z Liu, Y Xiong, Y Xiong, C Nie, H Gao, W Liu, X Ma (2021). Comparative and phylogenetic analysis of complete chloroplast genome sequences of two wildrye grasses *Elymus sibiricus* and *E. nutans* (Triticeae, Poaceae). *Intl J Agric Biol* 25:1203–1212

a small single-copy region (SSC) and two inverted repeat (IR) regions (Xiong *et al.* 2020; Hong *et al.* 2020). The cp genome of angiosperms is always 115–165 kb in length and contains about 130 genes, which are involved in photosynthesis, proteins encoding and transcriptional regulation (Daniell *et al.* 2016). Cp genome is not only necessary for the plant photosystem to promote photosynthesis and biomass yield, but also important in phylogenetic analysis and genetic diversity investigation due to its maternally inherited character and highly conserved genome sequences (Burke *et al.* 2012; Daniell *et al.* 2016; Zhang *et al.* 2016a). In particular, cp genome of plants is extraordinarily inherited from matrilineal line without interference of gene recombination, so its evolutionary path is correspondingly independent compared to the nuclear DNA (Ravi *et al.* 2008; Liu *et al.* 2018). The whole cp genome sequences and comparative analysis of some Triticeae species including genus *Triticum*, *Aegilops*, *Pseudoroegneria*, *Hordeum*, *etc.* (Gornicki *et al*. 2014; Chen *et al*. 2020). In addition, a point of concern is that *Pseudoroegneria* species with St-genome are generally considered the most likely maternal donor to *Elymus* L. genus, including *E. sibiricus* and *E. nutans* (Zuo *et al.* 2015). In this case, the comparative cp genome analysis between matrilineal lines and progeny species of *Elymus* would reveal their true evolutionary relationships.

The in-depth analysis of cp genome sequence for *E. sibiricus* and *E. mutans* is necessary to better understand the cp variance genes, structural variation of cp genomes and evolutionary relationships among *Elymus* species. Here, we present the *de* novo assembly and annotation of the cp genome sequence of *E. sibiricus* and *E. mutans*, and conduct a comparative analysis in order to (i) reveal the cp genomes variations between the two *Elymus* species and (ii) clarify the phylogenetic relationships between maternal species and progenies.

Materials and Methods

Plant Material, DNA extracting, and chloroplast genome sequencing

Fresh leaves of *E. sibiricus* (cv. Chuancao No.1) and *E. nutans* (cv. Aba) were collected in the country of Hongyuan, Aba Prefecture, Sichuan Province of China, located in southeastern Tibetan Plateau. Total DNA of one individual plant of each *Elymus* species were extracted using the Plant DNA Isolation Kit (ThermoFisher, Shanghai, China). Library construction and library quality testing were carried out after DNA quality was verified by 0.8% agarose gel. The chloroplast genomes were sequenced using Illumina Novaseq PE150 platform. The SPAdes v. 3.10.1 (Safonova *et al.* 2015) and Gapfiller v. 2.1.1 (Boetzer and Pirovano 2012) software were used to assemble the two studied *Elymus* cp genomes based on the reference cp sequence of *Hordeum vulgare* subsp. *vulgare* (KT 962228.1) retrieved from NCBI database. In addition, the cp genome coding sequences (CDS) were compared against by Blast v. 2.2.25 (Kent 2002) pipeline and used for gene annotation. The rRNA and tRNA gene sequences of chloroplast genomes were compared and predicted by Hmmer v. 3.1b2 software (Finn *et al.* 2011) and Aragorn v. 1.2.38 software (Laslett and Canback 2004), respectively. Lastly, Organellar Genome DRAW 1.3.1 (Lohse *et al.* 2013) was used to draw the circular cp genome map of *Elymus*.

Alignments analysis of multiple chloroplast genomes

The LAGAN mode of mVISTA (Poliakov *et al.* 2014) program was used to do multiple alignments of cp genomes between the two studied *Elymus* species, three *Pseudoroegneria* species, and *Hordeum vulgare* subspp. *Vulgare*, with *E.sibiricus* as reference. Homology and rearrangement occurrence of these species were analyzed in Mauve (Darling *et al.* 2010). Furthermore, the IRscope (Amiryousefi *et al.* 2018) online software was used to compare the boundary in the concatenation of IR and SC regions of the two *Elymus*, two *Pseudoroegneria* and *Hordeum vulgare* subsp. *vulgare* cp genomes.

Recognition of repetitive sequences

MISA v. 1.0 (Beier *et al.* 2017) software was used for extraction and recognition of the chloroplast Simple Sequence Repeats (cpSSRs). In addition, the inverted repetition (palindromic), direct repetition (forward), complement and reverse repetition with a minimum repetition length of 15 bp and sequence consistency greater than 90% were searched by REPuter 3.0 software (Kurtz *et al.* 2001).

Analysis of relative synonymous codon usage (RSCU)

The MEGA v. 7.0 software was used to analyze the RSCU, which reflect the relative preference of specific bases encoding the corresponding amino acid codons (Kumar *et al.* 2016). Values of RSCU greater than one was considered as better codon usage frequency.

Phylogenetic analysis and divergence time estimates

Cp genome sequences of studied *Elymus* and fourteen Triticeae published species in the NCBI database were conducted for phylogenetic analysis using BEAST v. 1.7.3 package; with three *Poaceae* cp genomes as outgroup. The GenBank numbers of the relative species are listed in Table S1.

Results

Chloroplast genomes features of *E. sibiricus* and *E. nutans*

The cp genome of *E. sibiricus* and *E. nutans* were

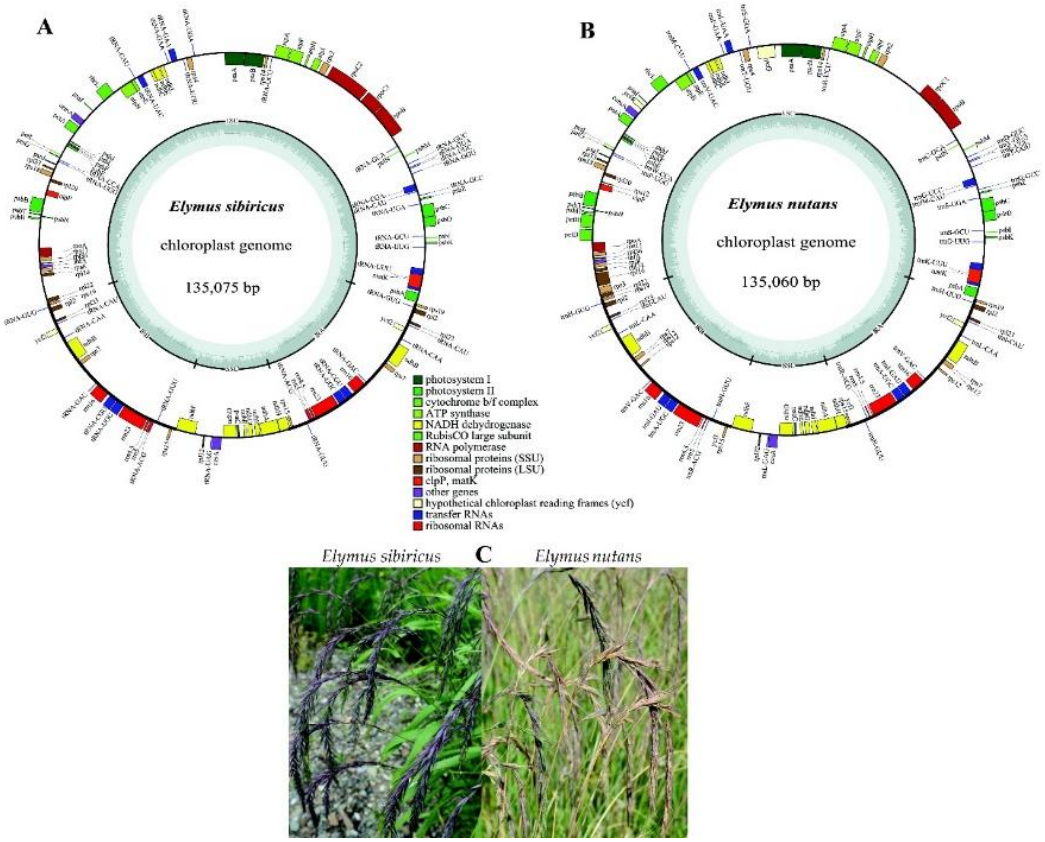

Fig. 1: Gene circle maps of the *Elymus sibiricus* (**A**) and *Elymus nutans* (**B**) chloroplast genomes. Genes belonging to different functional groups are color-coded. Genes transcribed clockwise and counterclockwise are indicated on the outside and inside of the large circle, respectively. The darker gray in the inner circle corresponds to GC content, whereas the lighter gray corresponds to AT content. Panicle morphology of two *Elymus* species (**C**)

sequenced and *de* novo assembled using Illumina short reads produced by genome skimming. The whole cp genome of *E. sibiricus* and *E. nutans* is 135,075 bp and 135,060 bp, respectively. Both their genomes have a typical quadripartite structure. Their cp DNA were divided into a LSC region of 80,681bp and 80,658 bp, an SSC region of 12,768 bp and 12,766 bp and two IR regions of 20,813 bp and 20,818 bp (Fig. 1 and Table 1). The guanine and cytosine (GC) contents of the cp genomes appeared very similar between *E. sibiricus* (38.34%) and *E. nutan* (38.33%).

The complete cp genomes of *E. sibiricus* and *E. nutans* contained 102 and 109 genes, respectively. Both cp genomes had four rRNAs, except that number of both tRNA and mRNA in *E. sibiricus* cp genome was slightly lower than *E. nutans* (Table 1). A total of 20 and 21 duplicated genes were found in IR of *E. sibiricus* and *E. nutans*, respectively. Two tRNA genes (*tRNA-CGA* and *tRNA-CGU*) only existed in *E. sibiricus* and three tRNA genes (*trnI-CAU, trnG-UCC* and *trnI-GAU*) were only found in *E. nutans* (Table 2 and 3). In the 47 photosynthesis-related genes, four genes (*ycf3, ycf4, petB and petD*) were unique to *E. nutans.* Among genes associated with encoding ribosomal proteins and transcription, *rps3, rps12, and rpl16* were specific in *E. nutans* while *rpl32* and *rpoC2* were unique to *E. sibiricus.* Additionally, only one pseudogenized *ycf1* gene was found in *E. nutans* (Table 3).

SSRs (simple sequence repeats) and interspersed repetitive sequences analysis

Interspersed repetitive (IR) sequences include palindrome repeats (P) and direct repeats (D). A total of 228 IR sequences were detected in the cp genome of *E. nutans*, which was higher than that of *E. sibiricus* (211). The percentage of type P repeats (48.25%, Fig. 2) in *E. nutans* was slightly lower than *E. sibiricus* (49.25%), but the type D repeats in *E. nutans* (51.75%) was slightly higher than *E. sibiricus* (50.7%).

A total of 165 and 161 SSRs were detected in cp genome of *E. sibiricus* and *E. nutans*, respectively. The single-bases A and T have the greatest number of repeat motifs in the two *Elymus* species (Fig. S1). A percentage of 77.0, 9.7 and 13.3% of SSRs were detected in LSC, SSC, and IR region of *E. sibiricus* (Fig. 3B). A very similar percentage pattern was found in *E. nutans* (Fig. 3C). In total,

Table 1: Comparison of the sequenced cp genomes of the two *Elymus* species

Sequence region	Length (bp)	
	E. sibiricus	*E. nutans*
Total cp genome	135075	135060
LSC region	80681	80658
SSC region	12768	12766
IR region	20813	20818
GC content	Percentage (%)	
Total cp genome	38.34	38.33
LSC region	36.37	36.37
SSC region	32.32	32.24
IR region	44	43.99
Gene Classification	Number	
Total genes	28	29
tRNA genes	4	4
rRNA genes	70	76
mRNA genes	102	109
Number of genes duplicated in IR	20	21

Table 2: Location and length of genes containing intron in two chloroplast genomes

Gene	*E. sibiricus*				*E. nutans*			
	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)	Location	Exon I (bp)	Intron I (bp)	Exon II (bp)
atpF	LSC	158	803	409	LSC	144	819	407
ndhA	SSC	550	1026	539	SSC	550	1026	539
ndhB	IRA	777	712	756	IRA	777	712	756
ndhB	IRB	777	712	756	IRB	777	712	756
tRNA-CGA	LSC	32	662	63	LSC	-	-	-
tRNA-CGU	IRA	32	787	59	IRA	-	-	-
tRNA-CGU	IRB	33	785	60	IRB	-	-	-
tRNA-UAA	LSC	36	575	51	LSC	35	574	50
tRNA-UAC	LSC	39	579	54	LSC	39	596	37
tRNA-UGC	IRA	37	811	36	IRA	37	811	35
tRNA-UGC	IRB	38	809	37	IRB	37	811	35
tRNA-UUU	LSC	39	2485	37	LSC	37	2488	35

Table 3: Comparison of the two *Elymus* species' chloroplast (cp) genomes

Category	Function	Name of genes
Self-replication (35)	Ribosomal RNA Genes	*rrn4.5, rrn5, rrn16, rrn23*
	Transfer RNA genes	*trnA-ACG, trnA-CAA, trnA-CAU, trnA-CCA, trnA-GAA, trnA-GAC, trnA-UAG, trnA-UAC*[*], *trnA-UUG, trnA-UUU*[*], *trnA-GCA, trnA-GCC, trnA-GGU, trnA-GUA, trnA-GUC, trnA-GUG, trnA-UGA, trnA-UGC*[*], *trnA-UGU, trnA-UGG, trnA-CGA*[*/es], *trnA-UCU, trnA-GCU, trnA-GUU, trnA-CGU*[*/es], *trnA-UUC, trnA-GGA, trnA-UAA*[*], *trnI-CAU*[en], *trnG-UCC*[*/en], *trnI-GAU*[*/en]
Ribosomal proteins (11) (translation)	Small subunit of ribosome (SSU)	*rps2, rps3*[en], *rps4, rps7, rps8, rps11, rps12*[*/en], *rps14, rps15, rps16, rps18, rps19*
Transcription (14)	Large subunit of ribosome (LSU)	*rpl2, rpl14, rpl16*[*/en], *rpl20, rpl22, rpl23, rpl32*[es], *rpl33, rpl36*
	RNA polymerase subunits	*rpoA, rpoB, rpoC1, rpoC2*[es]
	Translation initiation factor	*infA*
Photosynthesis related genes (47)	Large subunit of Rubisco	*rbcL*
	Subunits of Photosystem I	*psaA, psaB, psaC, psaI, psaJ, ycf3*[**/en], *ycf4*[en]
	Subunits of Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbT, psbZ, psbN*
	Subunits of ATP synthase	*atpA, atpB, atpE, atpF*[*], *atpH, atpI*
	Cytochrome b/f complex	*petA, petB*[*/en], *petD*[*/en], *petG, petL, petN*
	C-type cytochrome synthesis gene	*ccsA*
	Subunits of NADH dehydrogenase	*ndhA*[*], *ndhB*[*], *ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
Other genes (5)	Maturase	*matK*
	Protease	*clpP*
	Chloroplast envelope membrane protein	*cemA*
	Hypothetical protein	*ycf1*[en]
	Hypothetical open reading frames	*ycf2*

Note: Asterisk denotes the genes including a single intron; two asterisks denote the genes including two introns; [es], genes that are unique for *E. sibiricus;* [en], genes that are unique for *E. nutans*

71 SSRs existed in the exon region of *E. sibiricus*, while only 52 were found in the exon region of *E. nutans.* At the SSC region, there were six intergenic SSRs found only in *E. nutans* (Fig. 3A).

Feature of IR scope

The contraction and expansion of IR region were compared in the cp genomes of *E. sibiricus*, *E. nutans*,

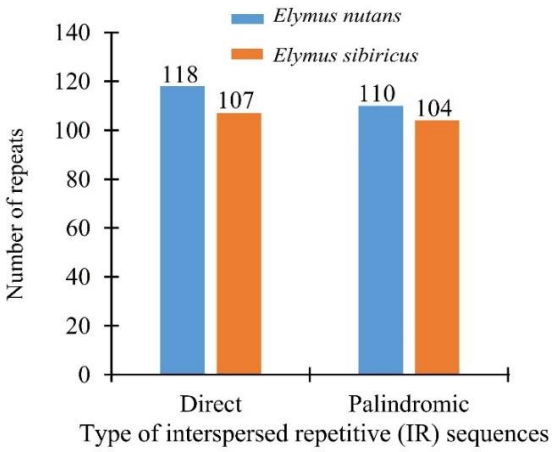

Fig. 2: Type and number distribution of repeat sequences in cp genomes of two *Elymus* species

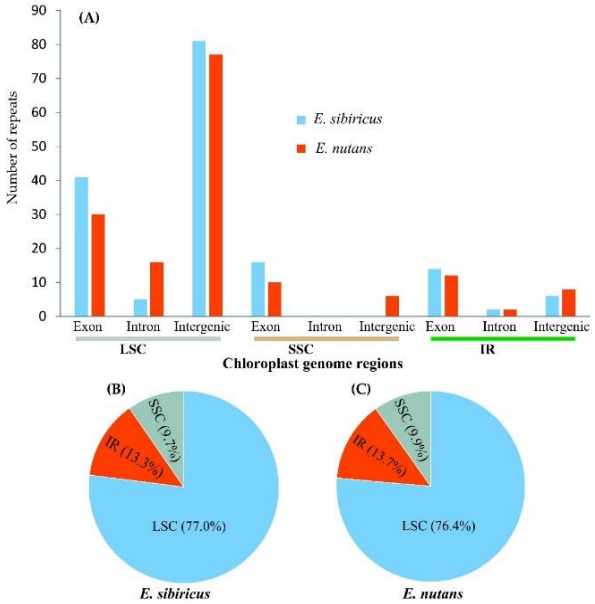

Fig. 3: Number (**A**) and frequency (**B**, **C**) of SSRs in the different region of *Elymus* cp genome

Pseudoroegneyia spicata, *P. libanotica* and *Hordeum vulgare* subsp. *vulgare* (Fig. 4). Overall, the result suggested little difference in the junction positions among the *Elymus* and *Pseudoroegneyia* cp genome sequences. There is a 34 bp spacer between *rpl22* genes and JBL (junction position of LSC and IRb region) in *E. sibiricus*, *P. spicata* and *P. libanotica*, whereas only 29 bp spacer was detected in *E. nutans.* Similarly, *rps19* gene and JLA (junction position of LSC and IRa region) were separated by a 48 bp spacer in *E. sibiricus*, *P. spicata* and *P. libanotica* and 53 bp spacer in *E. nutans.* More specifically, the gene *ycf1* was only detected in the IRa region of *E. nutans* and *P. libanotica.*

Variation analysis of six chloroplast genomes

The genetic variation among the two *Elymus* species, *Hordeum vulgare* ssp. *vulgare* and three *Pseudoroegneyia* cp genomes were analyzed via mVISTA (Poliakov *et al.* 2014) and Mauve (Darling *et al.* 2010). The results of the mVISTA revealed a lower variance in SSC and IR regions than in LSC regions, and more conservation in the coding regions than the non-coding regions (Fig. 5). The variation hotspot mainly existed in intragenic region. At the whole cp genome level, only a few variation hotspot regions existed in *Elymus* and *Pseudoroegneyia* species, which included *tRNA-GUC~psbM*, *tRNA-UAA~ndhJ*, *rbcL~psaI*, *rpl33~rps18*, and so on (Fig. 5). The result of the mVISTA analysis only between *E. sibiricus* and *E. nutans* shown that there were several hotspot regions (*tRNA-GUC~psbM*, *tRNA-UAA~ndhJ*, *rbcL~psaI*, *rpl33~rps18*, and so on). However, as shown in the local collinear block (Fig. S2), no inversion events or rearrangement were found among the six related species.

Analysis of relative synonymous codon usage

Relative synonymous codon usage (RSCU) is considered a combination of natural selection, genetic drift, and mutation. The RSCU of the two *Elymus* cp genomes was analyzed based on the 66 shared protein-coding genes (Fig. S3). We found that the RSCU values of initiation codon AUG were 1.991 and 1.982 in *E. sibiricus* and *E. nutans*, respectively. For three termination codons UAA, UAG, and UGA, the RSCU values were 1.6941, 0.6354 and 0.6705 in *E. sibiricus* and 1.7922, 0.5844 and 0.6234 in *E. nutans.* When the RSCU value of the codon was greater than one, it was considered a larger codon frequency. A 48.48% percentage (32 of 66, including three termination codons) of codons showed a greater frequency than one (RSCU > 1) both of two *Elymus* species, where 90.63% (29 of 32) codons prefer A+U at the third position.

Phylogenetic tree and divergence time

The Maximum-likelihood (ML) phylogenetic tree, based on the Bayesian MCMC (Markov Chain Monte Carlo) method, was obtained using the whole cp genome sequences of nineteen Poaceae species and, *Saccharum spontaneum*, *Sorghum bicolor*, and *Avena sativa* as outgroups (Fig. 6). Clearly, phylogenetic analysis supported the traditional phylogenetic classification of the Triticeae tribe. Two studied *Elymus* species and three *Pseudoroegneria* species were grouped in one clade, in which *E. sibiricus*, *E. nutans*, and three *Pseudoroegneria* species diverged around 3.061 Mya ago (Fig. 6). Approximately at 0.5746 Mya, *E. sibiricus*, *P. libanoticus* and *P. tauri* were divided, and around 0.4664 Mya the *E. nutans* and *P. spicata* were spitted from each other (Fig. 6), thus suggesting a close phylogenetic relationship between *E. nutans* and *P. spicata.*

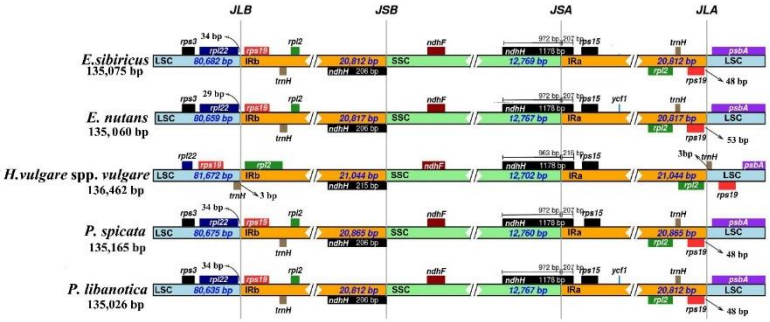

Fig. 4: IR scope analysis of cp genomes of five species. JLB, the junction position of LSC and IRb region; JSB, the junction position of SSC and IRb region; JSA, the junction position of SSC and IRa region; JLA, the junction position of LSC and IRa region

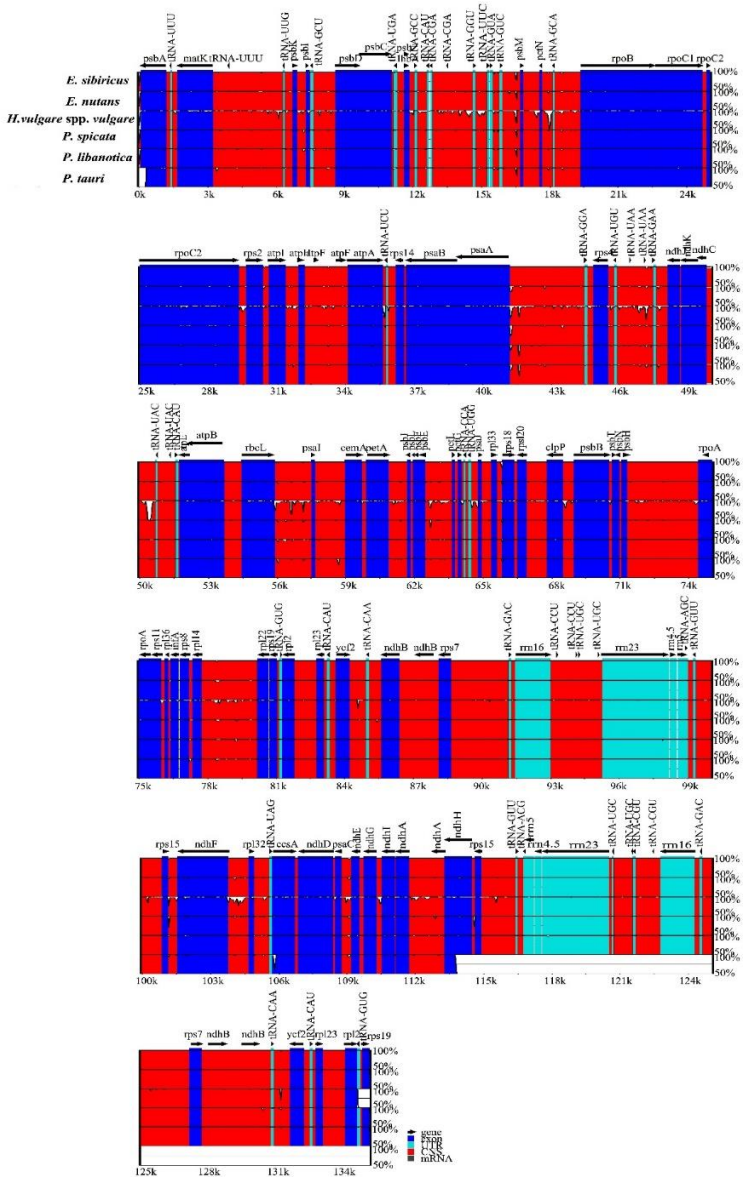

Fig. 5: Sequence identity plots among the two *Elymus* species and three *Pseudoroegneyia* species, with *E. sibiricus* as a reference. Annotated genes are shown on the top. Genome regions are color-marked as CNS (conserved non-coding sequences), exons, and introns. The color legend is summarized in the lower right-hand corner. Vertical scale indicates the percentage of identity ranging from 50% to 100%

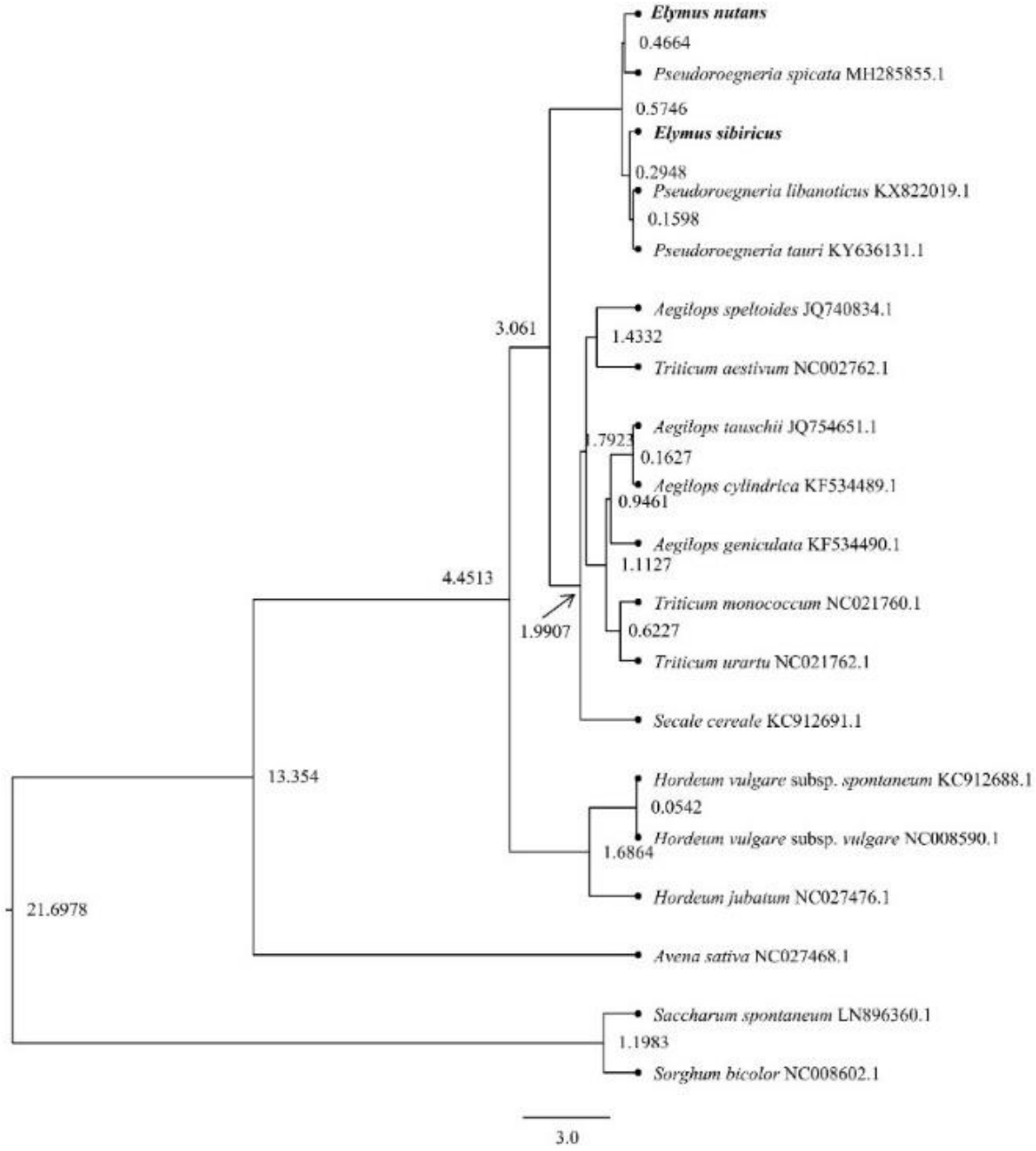

Fig. 6: Phylogenetic tree and divergence time among nineteen chloroplast genomes, the node value of the tree represents the average divergence time. The species used in this study are bolded

Discussion

Regularly, the 74 protein-coding genes were found in most angiosperms, while an additional five were found only in some species (Raman and Park 2016). However, 76 and 70 protein-coding genes were detected in *E. nutans* and *E. sibiricus*, respectively. These differential genes (*e.g.*, *ycf1*, *ycf3*, *ycf4*, *rps3*, *rps12*, *rpl32*) between the two *Elymus* species might be completely lost or transferred to the nuclear genome (Kan *et al.* 2020). In details, a unique pseudogenized *ycf1* gene was found only to exist in *E. nutans* and *Pseudoroegneyia libanotica*. The *ycf1* gene is functional and essential for cell survival in cp genomes of dicots except Poaceae (Huang *et al.* 2017). It is possible that the *ycf1* gene is not necessary for evolution and similarly to the *tufA* gene in angiosperms (Turmel *et al.* 2007), functionality of *ycf1* gene was transferred to the nuclear genome of those species that have one *ycf* pseudogene. Moreover, the *tRNA-CGA* in the LSC region and *tRNA-CGU, tRNA-CGU* in the IR region of *E. nutans* cp genome

have been lost. Although the cp genome of Poaceae is tremendously conservative, the subsistent differences will provide the basis for understanding the unique differences between related species or subspecies (Xiong *et al.* 2020).

Vast variant boundary regions of LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC are responsible for variations in cp genome size and rearrangement (Li *et al.* 2017). In addition, the *rpl22* gene, with the function of regulating senescence to maintain cell viability (Toro *et al.* 2019), showed a tendency of moving toward the IRb region in *E. nutans* compared with *E. sibiricus* and two *Pseudoroegneyia* species. It is well known that the most conservative quadripartite structure in the cp genome were the IR regions (Xiong *et al.* 2020). Therefore, this drift might help *rpl22* gene transfer into the IR region and further maintain the stability to attain the evolutionary adaptation of *E. nutans*.

CpSSR is one of the most significant tools to study genetic diversity, variety identification and phylogenetic analysis (Yamane and Kawahara 2018). Particularly, cp genomes have ancient patterns of inheritance that can offer insights into the evolutionary process (Cremen *et al.* 2018). Thus, the difference of cpSSRs in two *Elymus* could be used to further identify intraspecific genetic polymorphism. Except for cpSSR, many different cp DNA fragments and hotspot mutations could be used to develop barcode markers for congeneric species. There were many scattered mutational events existing in the cp genomes, which were generally gathered in "hotspots" and leaded a high variation region to distinguish the related species (Chao *et al.* 2017). In the two *Elymus* species we identified several hotspots regions, among them *tRNA-GUC~psbM*, *tRNA-UAA~ndhJ*, *rbcL~psaI*, *rpl33~rps18*, which could be used as new potential markers for future phylogenetic and phylo-geographic studies of *Elymus* species if available. Among these highly variable regions, the region of *rpl33~rps18* has been used as DNA barcodes in some plant species (Mariotti *et al.* 2010).

The RSCU values were calculated using the common genes of two *Elymus* species. The codon of leucine revealed the highest frequency (RSCU > 2), whereas the lowest frequency was found in the codon of methionine (RSCU < 0.02). The result was consistent with previous studies on cp genome of angiosperms (Li *et al*. 2019). Additionally, in agreement with Premal (Shah and Gilchrist 2011), we found that almost all of the codons with a high RSCU (RSCU > 1) value were A/U ended.

Chloroplast genome plays an important role in the evolutionary study due to the conservation of maternal inheritance (Nielsen *et al.* 2013). In this study, to obtain a more accurate evolutionary relationship and divergence time between *E. sibiricus* and *E. nutans*, their whole cp genome and other fifteen related species were used. The result showed that the *Elymus* and *Pseudoroegneyia* species separated from other Triticeae species about 3.061 million years ago (Mya). However, it is interesting to note that the *E. sibiricus* and *E. nutans* were grouped in two separate branches (Fig. 6). According to Chen *et al*. 2020, the shared St genome of *E. sibiricus* and *E. nutans* was both inherited from the *Pseudoroegneyia* species, while the respective specific species is not unambiguous. The phylogenetic relationships obtained in this study indicates that *E. natans* was more closely related to *Pseudoroegneyia spicata*, while *E. sibiricus* was closely related to *P. libanoticus* and *P. tauri*. Here, we could get a preliminary suggestion that the St nuclear genome of *E. sibiricus* originate from *P. libanoticus* or *P. tauri* and the St genome of *E. natans* originate from *P. spicata*. Of course, more evidence from nuclear genomes is required to support this view.

Conclusion

In present study, sequencing and *de* novo assembly of chloroplast genomes of *E. sibiricus* and *E. nutans* (Poaceae, Triticeae) were conducted using Illumina sequencing platform, which is an advantageous tool to research the origin and evolution of *Elymus* genus. We found that the structural characteristics of the two *Elymus* species have typical four-part structure in relationships similar to other Poaceae species. Large differences of interspersed repetitive sequences were detected between the two *Elymus* species. In addition, several hotspots (*e.g*., *tRNA-GUC~psbM*, *tRNA-UAA~ndhJ*, *rbcL~psaI*, *rpl33~rps18*) could be used to develop barcode marker for *Elymus* species. Finally, the phylogenetic analysis was in accordance with the traditional phylogenetic classification of the Triticeae tribe. This study provided new plastome insights into phylogenetic status and valuable gene resource in *Elymus* genus of Triticeae tribe.

Acknowledgments

We are very grateful to the Department of Grassland Science, Sichuan Agricultural University for providing us with experimental equipment and venue. Thanks to all authors for their hard work on this manuscript.

Funding

This work was supported by the Open project of the Key Laboratory of Utilization of Excellent Forage Germplasm Resources in Qinghai-Tibet Plateau of Qinghai Province (2020-ZJ-Y03) and Sichuan Province College Students' innovation and entrepreneurship training program (2019008006).

Grant Disclosures

The grant information was disclosed by the authors as "Open project of the Key Laboratory of Utilization of Excellent Forage Germplasm Resources in Qinghai-Tibet Plateau of Qinghai Province: 2020-ZJ-Y03. Sichuan Province College Students' innovation and entrepreneurship training program: 2019008006"

Author Contributions

Conceptualization, Xiao Ma; Methodology, Qingqing Yu and Zhechuan Liu; Resources, Wenhui Liu and Haidong Gao; Software, Yi Xiong, Zhechuan Liu and Yanli Xiong; Writing – original draft, Qingqing Yu and Xiao Ma; Writing – review & editing, Qingqing Yu, Yi Xiong, Cong Nie and Wenhui Liu.

Conflict of Interest

The authors declare there are no competing interests.

Data Availability

The annotated chloroplast genomes of *Elymus sibiricus* and *E. nutans* have been deposited in the NCBI GenBank with the accession numbers MT610375 and MT610376.

Ethics Approval

Not applicable.

References

Amiryousefi A, J Hyvonen, P Poczai (2018). IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34:3030–3031

Beier S, T Thiel, T Münch, U Scholz, M Mascher (2017). MISA-web: A web server for microsatellite prediction. *Bioinformatics* 16:2583–2585

Boetzer M, W Pirovano (2012). Toward almost closed genomes with GapFiller. *Genom Biol* 13:1–9

Burke SV, CP Grennan, MR Duvall (2012). Plastome sequences of two New World bamboos-*Arundinaria gigantea* and *Cryptochloa strictiflora* (Poaceae)-extend phylogenomic understanding of Bambusoideae. *Amer J Bot* 99:1951–1961

Chao X, D Wenpan, L Wenqing, L Yizeng, X Xiaoman, J Xiaobai, S Jipu, H Kaihong, S Zhili (2017). Comparative analysis of six *Lagerstroemia* complete chloroplast genomes. *Front Plant Sci* 8; Article 15

Chen N, W Chen, H Yan, Y Wang, H Kang, H Zhang, Y Zhou, G Sun, L Sha, X Fan (2020). Evolutionary patterns of plastome uncover diploid-polyploid maternal relationships in Triticeae. *Mol Phylogenet Evol* 149:106838

Cremen MCM, F Leliaert, VR Marcelino, H Verbruggen (2018). Large diversity of nonstandard genes and dynamic evolution of chloroplast genomes in Siphonous Green Algae (Bryopsidales, Chlorophyta). *Genom Biol Evol* 10:1048–1061

Daniell H, C Lin, M Yu, W Chang (2016). Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol* 17:134-162

Darling AE, B Mau, NT Perna (2010). ProgressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5; Article e11147

Finn RD, J Clements, SR Eddy (2011). HMMER web server: Interactive sequence similarity searching. *Nucl Acids Res* 39:29–37

Gornicki P, H Zhu, J Wang, GS Challa, Z Zhang, BS Gill, W Li (2014). The chloroplast view of the evolution of polyploid wheat. *New Phytol* 204:704–714

Hong Z, Z Wu, K Zhao, Z Yang, N Zhang, J Guo, LR Tembrock, D Xu (2020). Comparative analyses of five complete chloroplast genomes from the genus *Pterocarpus* (Fabacaeae). *Intl J Mol Sci* 21:3758-3775

Huang YY, ST Cho, M Haryono, CH Kuo (2017). Complete chloroplast genome sequence of common bermudagrass (*Cynodon dactylon* (L.) Pers.) and comparative analysis within the family Poaceae. *PLoS One* 12; Article e0179055

Kan S, T Shen, P Gong, J Ran, X Wang (2020). The complete mitochondrial genome of *Taxus cuspidata* (Taxaceae): Eight protein-coding genes have transferred to the nuclear genome. *BMC Evol Biol* 20; Article 10

Kent WJ (2002). BLAT—The BLAST-like alignment tool. *Genomics Res* 12:656–664

Kumar S, G Stecher, K Tamura (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874

Kurtz S, JV Choudhuri, E Ohlebusch, C Schleiermacher, J Stoye, R Giegerich (2001). REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucl Acids Res* 29:4633–4642

Laslett D, B Canback (2004). ARAGORN, a program to detect *tRNA* genes and *tmRNA* genes in nucleotide sequences. *Nucl Acids Res* 32:11–16

Lei Y, Y Zhao, F Yu, Y Li, Q Dou (2014). Development and characterization of 53 polymorphic genomic-SSR markers in Siberian wildrye (*Elymus sibiricus* L.). *Conserv Genet Resour* 6:861–864

Li Y, SP Sylvester, M Li, C Zhang, X Li, Y Duan, X Wang (2019). The complete plastid genome of *Magnolia zenii* and genetic comparison to Magnoliaceae species. *Molecules* 24:261-276

Li Y, Z Hansheng, P Zhenhua, L Dong, Z Gao (2014). Development and application of SSR molecular markers from the chloroplast genome of bamboo. *J Trop Subtrop Bot* 3:263–269

Li Z, H Long, L Zhang, Z Liu, H Cao, M Shi, X Tan (2017). The complete chloroplast genome sequence of tung tree (*Vernicia fordii*): Organization and phylogenetic relationships with other angiosperms. *Sci Rep* 7; Article 1869

Liu L, Y Wang, P He, P Li, J Lee, DE Soltis, C Fu (2018). Chloroplast genome analyses and genomic resource development for epilithic sister genera *Oresitrophe* and *Mukdenia* (Saxifragaceae), using genome skimming data. *BMC Genomics* 19; Article 235

Lohse M, O Drechsel, S Kahlau, R Bock (2013). Organellar Genome DRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucl Acids Res* 41:575–581

Mariotti R, NGM Cultrera, CM Diez, L Baldoni, A Rubini (2010). Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. *BMC Plant Biol* 10; Article 211

Nielsen AZ, B Ziersen, K Jensen, LM Lassen, CE Olsen, BL Moller, PE Jensen (2013). Redirecting photosynthetic reducing power toward bioactive natural product synthesis. *ACS Synth Biol* 2:308–315

Poliakov A, J Foong, M Brudno, I Dubchak (2014). GenomeVISTA—an integrated software package for whole-genome alignment and visualization. *Bioinformatics* 30:2654–2655

Qiao AH, JG Han, AQ Gong, W Li, YW Wang, GJ Qin, SD Guo, JM Wu, DZ Zhao (2006). Effect of nitrogen fertilizer application on *Elymus Nutans* seed quality and yield in Qinghai-Tibet plateau. *Acta Agrest Sin* 14:48–51

Raman G, S Park (2016). The complete chloroplast genome sequence of Ampelopsis: Gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. *Front Plant Sci* 7; Article 341

Ravi V, JP Khurana, AK Tyagi, P Khurana (2008). An update on chloroplast genomes. *Plant Syst Evol* 271:101–122

Safonova Y, A Bankevich, PA Pevzner (2015). dipSPAdes: Assembler for highly polymorphic diploid genomes. *J Comput Biol* 22:528–545

Shah P, MA Gilchrist (2011). Explaining complex codon usage patterns with selection for translational efficiency, mutation bias and genetic drift. *Proc Natl Acad Sci USA* 108:10231–10236

Toro ND, A Fernandez-Ruiz, L Mignacca, P Kalegari, MC Rowell, S Igelmann, E Saint-Germain, M Benfdil, S Lopes-Paciencia, L Brakier-Gingras (2019). Ribosomal protein *RPL22/eL22* regulates the cell cycle by acting as an inhibitor of the CDK4-cyclin D complex. *Cell Cycl* 18:759–770

Turmel M, JF Pombert, P Charlebois, C Otis, C Lemieux (2007). The green algal ancestry of land plants as revealed by the chloroplast genome. *Intl J Plant Sci* 168:679–689

Xiong Y, Y Xiong, S Jia, X Ma (2020). The complete chloroplast genome sequencing and comparative analysis of reed canary grass (*Phalaris arundinacea*) and harding grass (*P. aquatica*). *Plants* 9; Article 748

Yamane K, T Kawahara (2018). Size homoplasy and mutational behavior of chloroplast simple sequence repeats (cpSSRs) inferred from intra- and interspecific variations in four chloroplast regions of diploid and polyploid *Triticum* and *Aegilops* species. *Genet Resour Crop Evol* 65:727–743

Zhang D, K Li, J Gao, Y Liu, LZ Gao (2016a). The complete plastid genome sequence of the wild rice *Zizania latifolia* and comparative chloroplast genomics of the rice tribe Oryzeae, Poaceae. *Front Ecol Evol* 4; Article 88

Zhang Z, J Zhang, X Zhao, W Xie, Y Wang (2016b). Assessing and broadening genetic diversity of *Elymus sibiricus* germplasm for the improvement of seed shattering. *Molecules* 21; Article 869

Zhang Z, W Xie, J Zhang, N Wang, Y Zhao, Y Wang, S Bai (2019). Construction of the first high-density genetic linkage map and identification of seed yield-related QTLs and candidate genes in *Elymus sibiricus*, an important forage grass in Qinghai-Tibet Plateau. *BMC Genomics* 20; Article 861

Zuo H, P Wu, D Wu, G Sun (2015). Origin and reticulate evolutionary process of wheatgrass *Elymus trachycaulus* (Triticeae: Poaceae). *PLoS One* 10; Article e0125417