*Article*

Comparative analysis of complete chloroplast genome sequences of two wildrye grasses *Elymus* *sibiricus* and *E. nutans* (Triticeae, Poaceae)

Qingqing Yu 1, †, Yi Xiong 1, †, Zhechuan Liu1, †, Yanli Xiong 1, Cong Nie 1, Wenhui Liu 2, \*and Xiao Ma 1, \*

1 College of Animal Science and Technology, Sichuan Agricultural University, Chengdu 611130, China; yuqinggzu93@126.com (Q.Y.); xiongyi95@126.com (Y.X.); kangji61@126.com (Z.L.); yanlimaster@126.com (Y.X.); niecong125@163.com (C.N.)

2 Qinghai Academy of Animal Science and Veterinary Medicine, Key Laboratory of Superior Forage Germplasm in the Qinghai-Tibetan Plateau, Xi-ning, China; qhliuwenhui@163.com (W.L.)

\* Correspondence: maroar@126.com (X.M.); qhliuwenhui@163.com (W.L.); Tel./Fax: +86-028-86291010 (X.M.)

† These authors contributed equally to this work.

**Abstract**

Given their excellent characteristics like stress resistance and high biomass, *Elymus sibiricus* (containing the StH genomes) and *E. nutans* (containing the StHY genomes) are ecologically important species of *Elymus* genus in the tribe Triticeae (Poaceae). Theyare bothadaptive to grassland restoration in the eastern Tibetan Plateau. In this study, the complete chloroplast (cp) genome of *E. sibiricus* and *E. nutans* were sequencedand annotated with de novo analysis, to clarify their inter-species variation and their evolutionary relationships. The result showed that both two whole cp genomes contained a typical quadripartite structure and the cp genome length of *E. sibiricus* and *E. nutans* was 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 *Elymus* species. The phylogenetic analysis indicated that *E. sibiricus* and *E. nutans* diverged about 0.5746 million years ago (Mya). Comparative cp genome analysis of *E. sibiricus*, *E. nutans* and their reported St genome donors (*Pseudoroegneria*) were conducted, which provided the matrilineal evidence for their respective potential maternal ancestor species.

**Keywords:** *Elymus sibiricus*; *Elymus nutans;* 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 (LI Yong-xiang, 2005). 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 An-hai, 2006). Furthermore, the excellent genes resistance 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 and widely used perennial grasses species in the eastern Qinghai–Tibet Plateau (Zhang et al., 2016). *E. sibiricus* usually has a lower drought resistance and higher biomass yield than *E. nutans.* However, the high morphological similarity and niche overlaps limits their germplasm identification and further hinders seed production and promotion(Lei et al., 2014). The sequencing of the completed chloroplast (cp) genome via the next-generation sequencing technology (NGS) could provide a convenient and cost-effective approach to develop molecular markers, which are a potential tool for germplasm identification, evolutionary study, genetic relationship evaluation, and cp haplotype division (Li et al., 2014).

Chloroplast is an important component of plant organelles and photosynthetic organs (Douglas, 1994). The cp genome was reported to be consisted of a typical quadripartite structure with a large single-copy region (LSC), a small single-copy region (SSC) and two inverted repeat (IR) regions (Wicke et al., 2011; Douglas, 1994). 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 (Daniell et al., 2016; Zhang et al., 2016; Burke et al., 2012). 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 (Liu et al., 2018; Ravi et al., 2008; Allen, 2003). Previous study has indicated the common chloroplast haplotypes among species or subspecies using whole cp genome sequences of *Triticum* and *Aegilops* species (Gornicki et al., 2014). In addition, *Pseudoroegneria* species (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 the 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 *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 sequenced *Elymus* species and (ii) clarify the phylogenetic relationships between maternal species and progenies.

**Materials and methods**

**Material of plant, DNA extracting, and 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. The young leaves of two *Elymus* samples were snap-frozen in liquid nitrogen and stored at −80°C until DNA isolation. The flow cytometry has been used to measure nuclear DNA content in the leaves of two Elymus species and subsequently to allow species identification. Total DNA, which was used for next-generation sequencing, was obtained from fresh plant leaves with 0.1 g 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 Illumina Novaseq PE150 platform was then used to sequencing the total genomic DNA.

**Assembly and annotation of chloroplast genome**

The SPAdes v3.10.1 (Safonova et al., 2015) and Gapfiller v2.1.1 (Boetzer & 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) retrievedfrom NCBI database. Cp DNA sequence was first assembled using SPAdes software to obtain the seed sequence of cp genome. Kmer iterative extend seed was used to identify the pseudo contigs and merge them to acquire bracket though SSPACE v2.0 (Boetzer et al., 2010). Gapﬁller v2.1.1 procedure was used to calculate the GC content and create the gaps. In addition, the cp genome coding sequences (cds) in NCBI was used to compare against by Blast v2.2.25 (Kent, 2002) pipeline. The blast results were acquired and used for gene annotation in cp genome after artificial modification. The rRNA sequences of chloroplast genomes were compared by Hmmer v3.1b2 software (Finn et al., 2011). Furthermore, the tRNA sequence of cp genomes were predicted by Aragorn v1.2.38 software (Laslett & Canback, 2004). 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 (Frazer et al., 2004) program was used to do multiple alignments of cp genomes between the two studied *Elymus* species, three *Pseudoroegneria* species, and *Hordeum vulgare* subsp. *Vulgare*, with *E.sibiricus* as reference. In addition, the multiple alignments of cp genomes have been performed between *E.sibiricus* and *E.nutans.* Homology and rearrangement occurrence of these species were analyzed in Mauve (Wang et al., 2004). Furthermore, it is very important to evaluate the variation size of IR regions during evolution. 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 v1.0 (Beier et al., 2017) software was used for extraction and recognition of the chloroplast Simple Sequence Repeats (cpSSRs). In addition, the interspersed repetitive sequences, including 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%, which were searched throughout the cp genome by REPuter 3.0 (Kurtz et al., 2001) software.

**Analysis of relative synonymous codon usage**

Genome codon usage frequencies availability varies greatly among species. This unequal use of synonymous codons is known as RSCU, which was thought to be inﬂuenced by natural selection, mutation, and genetic drift. The MEGA v7.0 software was used to analyze the RSCU, which reﬂect the relative preference of specific bases encoding the corresponding amino acid codons (Kumar et al., 2008). Values of RSCU greater than one were considered as better codon usage frequency.

**Phylogenetic analysis and divergence time estimates**

Cp genomessequenceof studied *Elymus* and fourteen Triticeae published species in the NCBI database were collected to conduct the phylogenetic analysis, with three *Poaceae* cp genomes as outgroup. Firstly, MEGA v7.0 (Kumar et al., 2008) was used to blast the cp genome sequence, and then the BEAST v 1.7.3 package (Drummond et al., 2012) was used to calculate the divergence time through the alignment ﬁle with the Bayesian method. In order to conﬁrm that the value of eﬀective sample size (ESS) was greater than 200, the Tracer v 1.5 (<http://www.beast.bio.ed.ac.uk/>) was used to execute assessment of results. Finally, the Figtree v1.4.3 software was used to visualize the tree by the ﬁle obtained from TreeAnnotator. The GenBank numbers of the species we used were listed in Table S3.

**Results**

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

The cp genome of *E. sibiricus* and *E. nutans* were 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, a SSC region of 12,768 bp and 12,766 bp, and by two IR regions of 20,813 bp and 20,818 bp (Figure 1, Table 1). There were 38.34% and 38.33% guanine and cytosine (GC) contents of the cp genomes in the *E. sibiricus* and *E. nutans*, respectively.

The complete cp genomes of *E. sibiricus* and *E. nutans* contained 102 and 109 genes, respectively. Both cp genomes had four rRNAs: 29 tRNA and 76 mRNA were in the cp genome of *E. nutans*, and only 28 tRNA and 70 mRNA in the cp genome of *E. sibiricus* (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* (Tables 2, 3)*.* In the 47 photosynthesis-related genes, four genes (*ycf3, ycf4, petB and petD*) were unique to *E. nutans.* Furthermore, additional 11 genes possessed the function of encoding ribosomal proteins and 14 genes were associated with transcription. Among them, *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)*.*



Figure 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).

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

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Species | Size (bp) | | | | GC content (%) | | | | tRNA | rRNA | mRNA | genes | Number of genes duplicated in IR |
| cp | LSC | SSC | IR | cp | LSC | SSC | IR |
| *E. sibiricus* | 135075 | 80681 | 12768 | 20813 | 38.34 | 36.37 | 32.32 | 44.00 | 28 | 4 | 70 | 102 | 20 |
| *E. nutans* | 135060 | 80658 | 12766 | 20818 | 38.33 | 36.37 | 32.24 | 43.99 | 29 | 4 | 76 | 109 | 21 |

Table 2. Location and length of genes containing intron in the two *Elymus* 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* chloroplast (cp) genomes.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Category | Function | Name of gene | | | | | | | | | | |
| Self-replication (35) | Ribosomal RNA Genes | *rrn4.5* | *rrn5* | | *rrn16* | | *rrn23* | |  | | | |
| Transfer RNA genes | *trnA-ACG* | | *trnA-CAA* | | *trnA-CAU* | | *trnA-CCA* | | *trnA-CGA* *\*/*es | | *trnA-CGU* *\*/*es | |
| *trnA-GAA* | | *trnA-GAC* | | *trnA-UAG* | | *trnA-UAC\** | | *trnA-UCU* | | *trnA-UUC* | |
| *trnA-UUG* | | *trnA-UUU\** | | *trnA-GCA* | | *trnA-GCC* | | *trnA-GCU* | | *trnA-GGA* | |
| *trnA-GGU* | | *trnA-GUA* | | *trnA-GUC* | | *trnA-GUG* | | *trnA-GUU* | | *trnA-UAA\** | |
| *trnA-UGA* | | *trnA-UGC\** | | *trnA-UGU* | | *trnA-UGG* | | *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 asterisk denotes the genes including two introns; es, genes that are unique for *E. sibiricus;* en*,* genes that are unique for *E. nutans.*

**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 interspersed repetitive 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%, Figure 2) in *E. nutans* was slightly lower than *E. sibiricus* (49.25% in type P), but the type D repeats in *E. nutans* (51.75%) was slightly higher than *E. sibiricus* (50.7% in type D).



Figure 2. Type and number distribution of repeat sequences in the two *Elymus* cp genomes.

A total of 165 and 161 SSRs were detected in cp genome of *E. sibiricus* and *E. nutans,* respectively (Table S1). The single-bases A and T have the greatest number of repeat motifs in the two *Elymus* species (Figure S1). A percentage of 77.0%, 9.7% and 13.3% of SSRs were detected in LSC, SSC, and IR region of *E. sibiricus* (Figure 3B). A very similar percentage pattern was found in *E. nutans* (Figure 3C). 71 SSRs existed in the exon region of *E. sibiricus,* but only 52 exon SSRs were found in *E. nutans.* In the SSC region, there were only 16 exon SSRs in *E. sibiricus.* Moreover, 10 exon SSRs and six intergenic SSRs existed in *E.nutans* (Figure 3A).



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

**Feature of IR scope**

The contraction and expansion of IR region were compared in the cp genomes of *E. sibiricus*, *E. nutans*, *Pseudoroegneyia* *spicata*, *P.* *libanotica* and *Hordeum vulgare* subsp. *vulgare* (Figure 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* wasdetected in the IRa region of *E. nutans* and *P.libanotica.*



Figure 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.

**Variation analysis of six chloroplast genomes**

The genetic variation among the two *Elymus* species*,* *Hordeum vulgare* subsp. *vulgare* and three *Pseudoroegneyia* cp genomes were analyzed via mVISTA (Frazer et al., 2004) and Mauve (Darling et al., 2004). 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 (Figure 5). The variation hotspot mainly existed in intragenic region. At the whole cp genome level, only a few variation hotspot regions existed in *Elymus* species and *Pseudoroegneyia* species*,* which included *tRNA-GUC*～*psbM, tRNA-UAA*～*ndhJ, rbcL*～*psaI, rpl33*～*rps18,* and so on (Figure 5). The result of the mVISTA analysisbetween *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) (Figure S4). However, as shown in the local collinear block (Figure S2), no inversion events or rearrangement were found among the six related species.



Figure 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%.

**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 (Table S2 and Figure 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 is greater than one, it is considered a larger codon frequency. A 48.48% percentage (32 of 66, including three termination codons) of codons showed a codon frequency greater than one (RSCU > 1) in both in *E. sibiricus* and *E. nutans,* where 90.63% (29 of 32) codons prefer A+U at the third position.

**Phylogenetic tree and divergence time**

The Maximum-likelihood 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 (Figure 6). Clearly, phylogenetic analysis supported the traditional phylogenetic classification of the *Triticeae*.Two *Elymus* speciesand three *Pseudoroegneria* species were grouped in one clade, in which *E. sibiricus*, *E. nutans,* and three *Pseudoroegneria* species diverged around 3.061 Mya ago (Figure 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 (Figure 6); thus suggesting a close phylogenetic relationship between *E. nutans* and *P. spicata*.



Figure 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**

**The chloroplast genome feature of *Elymus sibiricus* and *E. nutans***

Regularly, the 74 protein-coding genes were found in most angiosperms, while the other five were found only in some species (Millen et al., 2001). 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* and so on) 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 chloroplast genomes of dicots except Poaceae (Steane, 2005). It is possible that the *ycf1* gene is not necessary for evolution and similarly to the *tufA* gene in angiosperms (Baldauf et al., 1990), 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 has been lost. Although the cp genome of Poaceae is tremendously conservative, the subsistent differences will provide the basis for understanding the unique differences between species or subspecies (Xiong et al., 2020).

The 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 (Del 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 and RSCU**

CpSSR is one of the most significant tools to study genetic diversity, variety identification, and phylogenetic analysis (Yamane & Kawahara, 2018). Particularly, cp genomes have ancient patterns of inheritance that can offer particular insights into the evolutionary process (Provan et al., 2001). 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 (Dong et al., 2012). In the two *Elymus* specieswe 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 (Figure S4). 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 were found in the codon of methionine (RSCU < 0.02). The result was consistent with previous studies on cp genome of angiosperms (Gu et al., 2004). Additionally, we found that almost all of the codons with a high RSCU (RSCU > 1) value were A/U ended; this result is also in agreement with Premal Shah et al. study (Shah & Gilchrist, 2011).

**Estimating phylogenetic and divergence time**

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), which is in accordance with the traditional classification. However, it is interesting to note that the *E. sibiricus* and *E. nutans* were grouped in two separate branches (Figure 6). According to Ning et al. (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 parent species is still unknown. 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 *Pseudoroegneyia libanoticus* and *Pseudoroegneyia* *tauri.* Here, we could get a preliminary suggestion that the St nuclear genome of *E. sibiricus* originate from *Pseudoroegneyia libanoticus* or *Pseudoroegneyia* *tauri*, and the St genome of *E. natans* originate from *Pseudoroegneyia* *spicata*. However, more evidence from nuclear genomes is required to support this suggestion.

**Conclusions**

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 similar typical four-part structure in relationships 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 insights into genetic evolution and valuable gene resource in *Elymus* genus.

**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.

**Additional information and declarations**

**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 following grant information was disclosed by the authors:

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

**Competing interests**

The authors declare there are no competing interests.

**Author contributions**

Conceptualization, Xiao Ma; Methodology, Qingqing Yu and Zhechuan Liu; Resources, Wenhui Liu; 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.

**Supplementary materials**

**Figure S1**: SSRs (Simple sequence repeats) of the two *Elymus* cp genomes. Types and number distribution of SSRs in the two *Elymus.***Figure S2**: Synteny comparison of four species chloroplast genomes using Mauve. Rectangular blocks of the same color indicate collinear regions of sequences. Vertical bars inside collinear blocks show degree of sequence identity. Dashed-line boxes indicate discovered plasmid sequence region.**Figure S3**: Relative synonymous codon usage of the two *Elymus* species. The characters in the box represent the codon subtype and the corresponding bar chart shows the frequency of codon usage.**Figure S4**: Sequence identity plots among the two *Elymus* species, with *Elymus nutans* 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. A cut-off of 50% identity was used for the plot. The vertical scale represents the percent identity between 50 and 100%. **Table S1**: Number and type of cpSSRs. **Table S2**: Relative synonymous codon usage. **Table S3**: GenBank number of each specie

**REFERENCES**

**Allen, J.F. 2003.** The function of genomes in bioenergetic organelles. *Philosophical Transactions of the Royal Society B* **358**:19-38. DOI: 10.1098/rstb.2002.1191.

**Amiryousefi, A., Hyvonen, J., and Poczai, P. 2018.** IRscope: an online program to visualize the junction sites of chloroplast genomes. *BIOINFORMATICS* **34**:3030-3031.

**Baldauf, S.L., Manhart, J.R., and Palmer, J.D. 1990.** Different fates of the chloroplast *tufA* gene following its transfer to the nucleus in green algae. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA* **87**:5317-5321. DOI: 10.1073/pnas.87.14.5317.

**Beier, S., Thiel, T., Munch, T., Scholz, U., and Mascher, M. 2017**. MISA-web: a web server for microsatellite prediction. *BIOINFORMATICS* **33**:2583-2585. DOI: 10.1093/bioinformatics/btx198.

**Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. 2010**. Scaffolding pre-assembled contigs using SSPACE. *BIOINFORMATICS* **27**:578-579. DOI: 10.1093/bioinformatics/btq683.

**Boetzer, M., and Pirovano, W. 2012**. Toward almost closed genomes with GapFiller. *GENOME BIOLOGY* **13**:1-9. DOI: 10.1186/gb-2012-13-6-r56.

**Burke, S.V., Grennan, C.P., and Duvall, M.R. 2012.** Plastome sequences of two New World *bamboos-Arundinaria gigantea* and *Cryptochloa strictiflora* (Poaceae)-extend phylogenomic understanding of Bambusoideae. *AMERICAN JOURNAL OF BOTANY* **99**:1951-1961. DOI: 10.3732/ajb.1200365.

**Chen, N., Chen, W., Yan, H., Wang, Y., Kang, H., Zhang, H., Zhou, Y., Sun, G., Sha, L., and Fan, X. 2020**. Evolutionary patterns of plastome uncover diploid-polyploid maternal relationships in Triticeae. *MOLECULAR PHYLOGENETICS AND EVOLUTION* **149**:106838. DOI: 10.1016/j.ympev.2020.106838.

**Daniell, H., Lin, C., Yu, M., and Chang, W. 2016**. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *GENOME BIOLOGY* **17**:134. DOI: 10.1186/s13059-016-1004-2.

**Darling, A.E., Craven, M., Mau, B., and Perna, N.T. 2004**. Multiple alignment of rearranged genomes. computational systems bioinformatics. p 738-739.

**Del Toro, N., Fernandez-Ruiz, A., Mignacca, L., Kalegari, P., Rowell, M.C., Igelmann, S., Saint-Germain, E., Benfdil, M., Lopes-Paciencia, S., and Brakier-Gingras, L. 2019**. Ribosomal protein RPL22/eL22 regulates the cell cycle by acting as an inhibitor of the CDK4-cyclin D complex. *CELL CYCLE*. DOI: 10.1080/15384101.2019.1593708.

**Dong, W., Liu, J., Yu, J., Wang, L., and Zhou, S. 2012**. Highly Variable Chloroplast Markers for Evaluating Plant Phylogeny at Low Taxonomic Levels and for DNA Barcoding. *PLoS One* **7**. DOI: 10.1371/journal.pone.0035071.

**Douglas, S.E. 1994**. *Chloroplast Origins and Evolution*. In: Bryant, D.A., ed. *The Molecular Biology of Cyanobacteria*. Springer Netherlands: Dordrecht, 91-118.

**Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. 2012**. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *MOLECULAR BIOLOGY AND EVOLUTION* **29**:1969-1973. DOI: 10.1093/molbev/mss075.

**Finn, R.D., Clements, J., and Eddy, S.R. 2011**. HMMER web server: interactive sequence similarity searching. *NUCLEIC ACIDS RESEARCH* **39**:29-37. DOI: 10.1093/nar/gkr367.

**Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. 2004**. VISTA: computational tools for comparative genomics. *NUCLEIC ACIDS RESEARCH* **32**:273-279. DOI: 10.1093/nar/gkh458.

**Gornicki, P., Zhu, H., Wang, J., Challa, G.S., Zhang, Z., Gill, B.S., and Li, W. 2014**. The chloroplast view of the evolution of polyploid wheat. *NEW PHYTOLOGIST* **204**:704-714. DOI: 10.1111/nph.12931.

**Gu, W., Zhou, T., Ma, J., Sun, X., and Lu, Z. 2004**. The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*. *BIOSYSTEMS* **73**:89-97. DOI: 10.1016/j.biosystems.2003.10.001.

**Kan, S., Shen, T., Gong, P., Ran, J., and Wang, X. 2020**. The complete mitochondrial genome of *Taxus cuspidata* (Taxaceae): eight protein-coding genes have transferred to the nuclear genome. *BMC EVOLUTIONARY BIOLOGY* **20**:10. DOI: 10.1186/s12862-020-1582-1.

**Kent, W.J. 2002**. BLAT—The BLAST-Like Alignment Tool. *GENOME RESEARCH* **12**:656-664. DOI: 10.1101/gr.229202.

**Kumar, S., Nei, M., Dudley, J.T., and Tamura, K. 2008**. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *BRIEFINGS IN BIOINFORMATICS* 9:299-306. DOI: 10.1093/bib/bbn017.

**Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. 2001**. REPuter: the manifold applications of repeat analysis on a genomic scale. *NUCLEIC ACIDS RESEARCH* 29:4633-4642. DOI: 10.1093/nar/29.22.4633.

**Laslett, D., and Canback, B. 2004**. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *NUCLEIC ACIDS RESEARCH* 32:11-16. DOI: 10.1093/nar/gkh152.

**Lei, Y., Zhao, Y., Yu, F., Li, Y., and Dou, Q. 2014**. Development and characterization of 53 polymorphic genomic-SSR markers in Siberian wildrye (*Elymus sibiricus* L.). *Conservation Genetics Resources* 6:861-864. DOI: 10.1007/s12686-014-0225-5.

**LI Yong-xiang, L.S.L.L. 2005**. Comparison of Genetic Diversity of Twelve *Elymus* Species Using ISSR and SSR Markers. *Scientia Agricultura Sinica* 8:1522-1527.

**Li, Y., Hansheng, Z., Zhenhua, P., Dong, L., and Gao, Z. 2014**. Development and Application of SSR Molecular Markers from the Chloroplast Genome of Bamboo. *Journal of Tropical and Subtropical Botany* 03:263-269.

**Li, Z., Long, H., Zhang, L., Liu, Z., Cao, H., Shi, M., and Tan, X. 2017**. The complete chloroplast genome sequence of tung tree (*Vernicia fordii*): Organization and phylogenetic relationships with other angiosperms. *Scientific Reports* 7:1869. DOI: 10.1038/s41598-017-02076-6.

**Liu, L., Wang, Y., He, P., Li, P., Lee, J., Soltis, D.E., and Fu, C. 2018.** Chloroplast Genome Analyses and Genomic Resource Development for Epilithic Sister Genera *Oresitrophe* and *Mukdenia* (Saxifragaceae), Using Genome Skimming Data. *BMC GENOMICS* 19:1-17. DOI: 10.1186/s12864-018-4633-x.

**Lohse, M., Drechsel, O., Kahlau, S., and Bock, R. 2013**. Organellar Genome DRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *NUCLEIC ACIDS RESEARCH* 41:575-581. DOI: 10.1093/nar/gkt289.

**Mariotti, R., Cultrera, N.G.M., Diez, C.M., Baldoni, L., and Rubini, A. 2010**. Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. *BMC PLANT BIOLOGY* 10:211. DOI: 10.1186/1471-2229-10-211.

**Millen, R.S., Olmstead, R.G., Adams, K.L., Palmer, J.D., Lao, N.T., Heggie, L., Kavanagh, T.A., Hibberd, J.M., Gray, J.C., and Morden, C.W. 2001**. Many Parallel Losses of *infA* from Chloroplast DNA during Angiosperm Evolution with Multiple Independent Transfers to the Nucleus. *The Plant Cell* 13:645-658. DOI: 10.1105/tpc.13.3.645.

**Nielsen, A.Z., Ziersen, B., Jensen, K., Lassen, L.M., Olsen, C.E., Moller, B.L., and Jensen, P.E. 2013.** Redirecting Photosynthetic Reducing Power toward Bioactive Natural Product Synthesis. *ACS Synthetic Biology* 2:308-315. DOI: 10.1021/sb300128r.

**Provan, J., Powell, W., and Hollingsworth, P.M. 2001**. Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends in Ecology and Evolution* 16:142-147. DOI: 10.1016/S0169-5347(00)02097-8.

**QIAO An-hai, H.J.G.A. 2006**. Effect of Nitrogen Fertilizer Application on *Elymus Nutans* Seed Quality and Yield in Qinghai-Tibet Plateau. p 48-52. DOI: 10.1360/aps040178.

**Ravi, V., Khurana, J.P., Tyagi, A.K., and Khurana, P. 2008**. An update on chloroplast genomes. *PLANT SYSTEMATICS AND EVOLUTION* 271:101-122. DOI: 10.1007/s00606-007-0608-0.

**Safonova, Y., Bankevich, A., and Pevzner, P.A. 2015**. dipSPAdes: Assembler for Highly Polymorphic Diploid Genomes. *JOURNAL OF COMPUTATIONAL BIOLOGY* 22:528-545. DOI: 10.1089/cmb.2014.0153.

**Shah, P., and Gilchrist, M.A. 2011.** Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc Natl Acad Sci U S A* 108:10231-10236. DOI: 10.1073/pnas.1016719108.

**Steane, D.A. 2005**. Complete Nucleotide Sequence of the Chloroplast Genome from the Tasmanian Blue Gum, *Eucalyptus globulus* (Myrtaceae). *DNA RESEARCH* 12:215-220. DOI: 10.1093/dnares/dsi006.

**Wang, T., Touchman, J.W., and Xue, G. 2004.** Proceedings. 2004 IEEE Computational Systems Bioinformatics Conference. Computational Systems Bioinformatics Conference. DOI: 10.1109/CSB.2004.1332396.

**Wicke, S., Schneeweiss, G.M., Depamphilis, C.W., Muller, K.F., and Quandt, D. 2011**. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *PLANT MOLECULAR BIOLOGY* 76:273-297. DOI: 10.1007/s11103-011-9762-4.

**Xiong, Y., Xiong, Y., Jia, S., and Ma, X. 2020**. The Complete Chloroplast Genome Sequencing and Comparative Analysis of Reed Canary Grass (*Phalaris arundinacea*) and Hardinggrass (*P. aquatica*). *Plants* 9:748. DOI: 10.3390/plants9060748.

**Yamane, K., and Kawahara, T. 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. *GENETIC RESOURCES AND CROP EVOLUTION* 65:727-743. DOI: 10.1007/s10722-017-0567-4.

**Zhang, D., Li, K., Gao, J., Liu, Y., and Gao, L. 2016.** The Complete Plastid Genome Sequence of the Wild Rice *Zizania latifolia* and Comparative Chloroplast Genomics of the Rice Tribe Oryzeae, Poaceae. *Frontiers in Ecology and Evolution* 4. DOI: 10.3389/fevo.2016.00088.

**Zhang, Z., Zhang, J., Zhao, X., Xie, W., and Wang, Y. 2016**. Assessing and Broadening Genetic Diversity of *Elymus sibiricus* Germplasm for the Improvement of Seed Shattering. *Molecules (Basel, Switzerland)* 21:869. DOI: 10.3390/molecules21070869.

**Zuo, H., Wu, P., Wu, D., and Sun, G. 2015**. Origin and Reticulate Evolutionary Process of Wheatgrass *Elymus trachycaulus* (Triticeae: Poaceae). *PLoS One* 10: e125417. DOI: 10.1371/journal.pone.0125417.