The complete chloroplast genomes of seventeen *Aegilops tauschii*: Genome comparative analysis and phylogenetic inference

Qing Su \(^1\), Luxian Liu \(^1\), Mengyu Zhao \(^1\), Cancan Zhang \(^1\), Dale Zhang \(^1\), Youyong Li \(^2\), Suoping Li \(^1\)

\(^1\) Key Laboratory of Plant Stress Biology, School of Life Science, Henan University, Kaifeng, China
\(^2\) School of Life Science and Technology, Henan Institute of Science and Technology, Xinxiang, China

Corresponding Authors: Youyong Li, Suoping Li
Email address: liyouyong@163.com, lisuoping@henu.edu.cn

The D genome progenitor of bread wheat, *Aegilops tauschii* Cosson (DD, 2n = 2x = 14), which is naturally distributed in areas of central Eurasia, ranging from northern Syria and Turkey to western China, is considered a potential genetic resource for improving bread wheat. In this study, the chloroplast (cp) genomes of 17 *Ae. tauschii* accessions were reconstructed. The cp genomes ranged from 135,551 bp to 136,009 bp and contained a typical quadripartite structure of angiosperms. Within these genomes, we identified a total of 127 functional genes, including 78 protein-coding genes, 26 transfer RNA genes and four ribosomal RNA genes, with 19 duplicated genes in the IRs. Although comparative analysis revealed that the genomic structure (gene order, gene number and IR/SC boundary regions) is conserved, a few variant loci were detected, predominantly in the non-coding regions (intergenic spacer regions). IR expansion and contraction was completely identical among 17 *Aegilops tauschii* accessions, which not influence the chloroplast genomes in length. The phylogenetic relationships determined based on the complete genome sequences were consistent with the hypothesis that *Ae. tauschii* populations in the Yellow River region of China originated in South Asia not Xinjiang province. Furthermore, we confirmed that *Ae. tauschii* was derived from monophyletic speciation rather than hybrid speciation at the chloroplast genome level. We also identified four variable genomic regions, *rpl32-trnL-UAG*, *ccsA-ndhD*, *rbcL-psal* and *rps18-rpl20*, showing high levels of nucleotide polymorphisms, which may accordingly prove useful as cpDNA markers in studying the interspecific and intraspecific genetic structure and diversity of *Ae. tauschii*. 
The complete chloroplast genomes of seventeen *Aegilops tauschii*: Genome comparative analysis and phylogentic inference

Qing Su¹⁺, Luxian Liu¹⁺, Mengyu Zhao¹, Cancan Zhang¹, Dale Zhang¹, Youyong Li²*, Suoping Li¹*

¹Key Laboratory of Plant Stress Biology, School of Life Science, Henan University, Kaifeng 475000, China
²School of Life Science and Technology, Henan Institute of Science and Technology, Xinxiang 453003, China

+ The two authors contributed equally to this work

* Corresponding author

Email list for all authors:

Qing Su: suqingangel@163.com; Luxian Liu: liushuangcx2007@126.com; Mengyu Zhao: 835591597@qq.com; Cancan Zhang: 894275874@qq.com; Dale Zhang: zhangdale@henu.edu.cn; Youyong Li: liyouyong@163.com; Suoping Li: lisuoping@henu.edu.cn
Abstract

The D genome progenitor of bread wheat, *Aegilops tauschii* Cosson (DD, 2n = 2x = 14), which is naturally distributed in areas of central Eurasia, ranging from northern Syria and Turkey to western China, is considered a potential genetic resource for improving bread wheat. In this study, the chloroplast (cp) genomes of 17 *Ae. tauschii* accessions were reconstructed. The cp genomes ranged from 135,551 bp to 136,009 bp and contained a typical quadripartite structure of angiosperms. Within these genomes, we identified a total of 122 functional genes, including 82 protein-coding genes, 32 transfer RNA genes and eight ribosomal RNA genes, with 16 duplicated genes in the IRs. Although comparative analysis revealed that the genomic structure (gene order, gene number and IR/SC boundary regions) is conserved, a few variant loci were detected, predominantly in the non-coding regions (intergenic spacer regions). The phylogenetic relationships determined based on the complete genome sequences were consistent with the hypothesis that *Ae. tauschii* populations in the Yellow River region of China originated in South Asia not Xinjiang province or Iran, which could be contribute to more effective utilization of wild germplasm resources. Furthermore, we confirmed that *Ae. tauschii* was derived from monophyletic speciation rather than hybrid speciation at the chloroplast genome level. We also identified four variable genomic regions, *rpl32-trnL-UAG, ccsA-ndhD, rbcL-psaI* and *rps18*...
rpL20, showing high levels of nucleotide polymorphisms, which may accordingly prove useful as cpDNA markers in studying the intraspecific genetic structure and diversity of *Ae. tauschii*.

**Keywords:** *Aegilops tauschii*, next-generation sequencing, phylogenetic tree, nucleotide polymorphisms, spreading route

**Introduction**

*Aegilops tauschii* Cosson (DD, 2n = 2x = 14), which is the D genome progenitor of common bread wheat, is characterized by abundant genetic variation and is noted for its strong tillering ability and high plant tolerance (resistance to disease, drought and abiotic stress) (Singh et al., 2012). *Ae. tauschii* is widely distributed in central Eurasia with a range that extends from northern Syria and Turkey to western China (the Yili area of Xinjiang). In addition, as a farmland weed that co-occurs with common wheat, *Ae. tauschii* is also found in the middle reaches of the Yellow River (including Henan and Shannxi provinces, China) (Wei et al., 2008). On the basis of the findings of genetic studies, *Ae. tauschii* has been divided into two sublineages based on nuclear genome sequences, namely L1 and L2, which are broadly affiliated with *Ae. tauschii* ssp. *strangulata* and *Ae. tauschii* ssp. *tauschii*, respectively (Mizuno et al., 2010; Dvorak et al., 1998). Previous studies have proved that L2 lineage is involved in the origin of common wheat and is limited to a narrow area within the whole species’ overall distribution range (Wang et al., 2013; Dvorak et al., 2012), as well as L1 lineage is adapted to more diverse environments (Dudnikov, 2012). Given the large genetic distance between L1 and L2, it has been proposed that *Ae. tauschii* (especially L1 lineage) shows more genetic differentiation than does the D genome of the common wheat (Lubbers et al., 1991; Wang et al., 2013; Dvorak et al., 2012). Thus, in common with wild crop progenitors, *Ae. tauschii*, especially L1 lineage, is considered to represent a potentially valuable germplasm resource that could be exploited for genetic improvement of common wheat (Kilian et al., 2011; Zhang et al., 2017; Zhang et al., 2018). Moreover, owing to widely geographical distribution of the L1 lineage, studying of its genetic and evolutionary relationship could be contribute to more effective utilization of wild germplasm resources.
Iran is widely regarded as the center of the origin and genetic diversity of *Ae. tauschii* (Dvorak et al., 1998). However, as a consequence of long periods of dispersal and adaptation, this species now shows distinct difference within its distribution range. At present, L2 lineage is mainly restricted to an area extending from Transcaucasia (Armenia and Azerbaijan) to eastern Caspian Iran, whereas L1 lineage is more widely distributed across the entire species range, which includes the middle reaches of the Yellow River and the Yili area of Xinjiang Province in China (Kihara et al., 1965; Jaaska 1980; Matsuoka et al., 2015; Wei et al., 2008). As a clade of L1 lineage, there is still some debate concerning the specific spreading route from Xinjiang to the middle reaches of the Yellow River in China, given the large geographical isolation of these two areas. Some researchers believe that the long-distance spread occurred concomitantly with the expansion of original wheat varieties and *Ae. tauschii* accessions collected from the Yellow River are introduced from Xinjiang province (Wang et al., 2010; Yen et al., 1984). However, on the basis of the established genetic similarities among 31 *Ae. tauschii* accessions from China and Iran determined using SSR markers, Wei et al., (2008) proposed that *Ae. tauschii* populations in the Yellow River region are likely to have been directly introduced from Iran along the silk road, which indicates a longer genetic relationship with Xinjiang landraces. In addition, Mizuno et al., (2010) used AFLP molecular markers to classify genetic types among *Ae. tauschii* accessions and found that the accessions collected in the Yellow River basin, which were significantly different from the L1E accessions in Afghanistan and Xinjiang, closely clustered to the L1W accessions in the Middle East.

With regard to the introduction of *Ae. tauschii* in China, it is believed that the routes by which barley and wheat spread from western Eurasian to East Asia may offer some relevant clues. One hypothesis postulated that common wheat arrived from the west (probably from Afghanistan or the Central Asia oases), moving eastwards into northern Xinjiang (Betts et al., 2014; Wu et al., 2019), whereas an alternative opinion envisaged that common wheat may have reached East Asia along a north-west passage from Eurasia, via southern Siberia and Mongolia. In both scenarios, common wheat was assumed to have initially spread into central Gansu and thereafter migrated eastwards into the Yellow River region or was later introduced to the margins of the Qinghai–Tibetan plateau in highland China (Dodson et al., 2013; Betts et al., 2014; Long et al., 2018; Wu et al., 2019). A subsequent dispersal route from the margins of the Qinghai–Tibetan plateau to the Yangtze valley has also been proposed (Wu et al., 2019). It is generally considered that barley had arrived on
the northeastern and southeastern Tibetan plateau at a date prior to 4000 calendar years before the present (Wu et al., 2019). However, in contrast to the aforementioned traditional views regarding the dispersal routes of common wheat and barley (Dodson et al., 2013; Betts et al., 2014; Long et al., 2018), Zeng et al (2018) have suggested that qingke barley is derived from eastern domesticated barley and was introduced from south Tibet, most likely from northern Pakistan, India, and Nepal eastwards into the Tibetan plateau, which is supported by recent archaeological evidence of the occurrence of barley in north-east India. Thus, the specific route whereby *Ae. tauschii* extended its range is still ambiguous, and remained considerable interest for intensive exploration.

Owing to the notable prominent advantages of the chloroplast genome, including a moderate rate of nucleotide replacement, significant variations in the rate of molecular evolution between non-coding and coding regions, moderate genome size, and desirable collinear properties among different species (Liu et al., 2018), analysis of chloroplast sequences is considered an effective strategy for investigating intra- and interspecific evolutionary relationships, as well as being of considerable utility in comparative genomic studies (Matsuoka et al., 2005; Yamane and Kawahara 2005; Tabidze et al., 2014; George et al., 2015; Liu et al., 2017; Shang et al., 2019). In angiosperms, the size of cp genomes and their gene arrangements are generally highly conserved with a circular chromosome ranging in size from 120 to 160 kb, and comprising a small single-copy (SSC) region, a large single-copy (LSC) region and a pair of inverted repeats (IRs) region (Palmer, 1991; Yang et al., 2010). Given that chloroplasts are primarily non-recombining and uniparentally inherited, phylogenetic analysis based on chloroplast sequences can also facilitate specific identification of maternal lineages (Sang, 2002).

Using complete cp genomes, a number of previous studies have examined genetic variations in common wheat and its relatives and performed related phylogenetic analyses (Middleton et al., 2014; Gornicki et al., 2014; Gogniahshvili et al., 2016). In the present study, with a view toward shedding light on the genetic variation of *Ae. tauschii* and the source of Chinese landraces, we performed sequence analysis of the complete cp genomes of 17 *Ae. tauschii* accessions derived from regions spanning the known distribution range from western Turkey to eastern China. The results not only provided a series of new insights to clarify the spreading route of *Ae. tauschii*, but also enabled us to identify promising germplasm resources for the genetic improvement of bread wheat.
Materials and methods

Plant materials

Information relating to the 17 *Ae. tauschii* accessions examined in this study, including origin and collection region, are listed in Table 1. The accessions marked as ‘XJ’, ‘T’, and ‘S’ are those from Xinjiang, Henan, and Shannxi provinces, respectively. These materials were collected in the field by Key Laboratory of Plant Stress Biology of Henan University. The seeds with the accessions designated ‘AY’ and ‘AS’ were provided by the US National Plant Germplasm Center and Institute of Genetics and Developmental Biology. The sources of these accessions are distributed across a wide geographical range that extents from Turkey, Georgia, Iran, Turkmenistan, Kazakhstan, Tajikistan, Afghanistan, Pakistan, and India to Xinjiang, Shannxi and Henan provinces in China (Figure 1), and includes 15 lineage L1 accessions and two lineage L2 accessions, as determined based on single-nucleotide polymorphisms (SNPs) (Wang et al., 2013).

Next-generation sequencing, annotation and comparison of chloroplast genomes

For each of the 17 *Ae. tauschii* accessions, total genomic DNA was extracted from fresh leaves of 1-week-old seedlings germinated by the aboved seeds of *Ae. tauschii* by the method of Han et al. (2015). Approximately 5 to 10 μg of extracted DNA was sheared to generate fragments, and the quality of DNA sequences was determined using an Agilent Bioanalyzer 2100 (Agilent Technologies). Thereafter, we generated a paired-end sequencing library, which was constructed from ~400 bp fragments obtained using a Genomic DNA Sample Prep Kit (Illumina) in accordance with the manufacturer’s protocol. Subsequent genome sequencing was carried out by Majorbio Bio-Pharm Technology Co., Ltd. (Shanghai, China) using the HiSeq X ten sequencing platform (Illumina Inc., San Diego, CA) with 150 bp read length. Low quality reads with a phred score < 30 and 0.001 probability error were removed using Trimmomatic v0.36 (http://www.usadellab.org/cms/index.php?page=trimmomatic), and the remaining high quality fragments were assembled into contigs using the SOAPdenovo v2.21 (http://soap.genomics.org.cn/soapdenovo.html). The assembled contigs were further aligned to the AL8/78 reference genome (GenBank No. KJ 614412) using BLASTN (http://www.ncbi.nlm.nih.gov). Finally, gaps in the genome sequence were filled using GapCloser v1.12 (http://soap.genomics.org.cn/soapdenovo.html).

The complete chloroplast genome of *Ae. tauschii* was annotated using the Dual Organellar GenoMe Annotator program (DOGMA, Wyman et al., 2004), and then verified artificially. The annotation of the tRNA genes was verified using tRNAscan-SE. The circular chloroplast genome of *Ae. tauschii* was constructed using OGDRAW online software (Lohse et al., 2013). The Gene Ontology (GO) functional categories were executed using Blast2GO V3.2 software.

In order to determine intraspecific variations, variation in the LSC and SSC region, and the expansion and
interspecific variations of IR regions, we used Geneious 9.0.5 software (Biomatters, Auckland, New Zealand). Using AL8/78 as the reference genome, the chloroplast genomes of T093 and AY81 were performed to visualize the sequence variations by mVISTA.

### Phylogenetic analysis

The phylogenetic relationships among *Ae. tauschii* accessions were examined using the complete chloroplast genome sequences of the 17 accessions. In addition, in order to establish the origin of *Ae. tauschii* within the family Triticeae, we performed phylogenetic analysis using the genome sequences of the following 99 accessions: the 17 accessions newly sequenced in the present study, 56 accessions in the genus *Aegilops*, 24 accessions in the genus *Triticum*, and two accessions of *Hordeum vulgare* as outgroups.

The sequences of these chloroplast genomes were aligned using Geneious 9.0.5 software (Biomatters, Auckland, New Zealand). Gaps were adjusted manually or removed. The alignment lengths with all gap positions removed were determined to be 135,984 bp (17 accessions) and 131,116 bp (99 accessions). We constructed two corresponding phylogenetic trees based on the maximum-likelihood (ML) method utilizing MEGA7.0 software (Kumar et al., 2016). The bootstrap values were evaluated with 1000 replications. The best-fit nucleotide substitution model (GTR+I+G) was selected using jModelTest 2.1.4 software (Posada, 2008) operating with default parameters.

### Analysis of sequences divergence and molecular markers

Analysis of chloroplast genome divergence among the 17 *Ae. tauschii* accessions was performed by employing MAFFT alignment based on homologous genomic regions (Katoh et al., 2002). Nucleotide diversity (Pi) of polymorphic sequences among 17 *Ae. tauschii* individuals were calculated utilizing DnaSP v5.0 software (Librado and Rozas, 2009).

### Results

#### The characteristic and comparison of *Ae. tauschii* chloroplast genomes

The chloroplast genomes of the 17 *Ae. tauschii* accessions sequences in the present study were found to be similar with respect to size and gene content. They are generally 135,551 ~ 136,009 bp in length and contained four quadripartite structure: two IR regions separated by a SSC region (12,773 ~ 12,826 bp) and a LSC region (79,723 ~ 80,140 bp) (Figure2 and Table 2). The intergenic regions of these sequences ranged within 75,635 ~ 77,699 bp in length, accounting for 53.2 to 57.2 % of the total genomes. The remaining sequence gene regions
comprise coding regions, intron regions, tRNA genes, and rRNA genes. Notably the GC contents of the 17 chloroplast genomes are relatively stable ranging from 38.31 % to 38.35 % (Table 2).

We also found the compositions of 17 genomes to be highly conserved, with each of the accessions containing a total of 122 functional genes (82 protein-coding genes, 8 rRNAs, and 32 tRNAs), of which 16 were duplicated genes (Table 2). Among the 106 unique genes, 56 are related to self-replication and 46 are associated with photosynthesis. The functions of the remaining four annotated genes are as follows: a maturase (\textit{matK}), an envelope membrane protein (\textit{cemA}), C-type cytochrome synthesis (\textit{ccsA}) as well as a protease (\textit{clpP}) (Table S1). As shown in Table S1, the 122 functional genes were were classified into three groups and assigned according annotations clearly. Then, the Gene Ontology (GO) functional categories among these genes were investigated which were mainly enriched for cellular process, metabolic process, cell, cell part and organelle using blast2go (Figure S1 and Table S2).

The sequences of representative genomes of three \textit{Ae. tauschii} accessions were relatively conserved, and any translocations or inversions among these genomes were not identified. The IR regions show lower sequence variation than that either of the LSC and SSC regions (Figure S2). Further, the precise IR/SSC and IR/LSC boundary locations and the corresponding neighboring genes were found to be identical in length (21,548 bp) and border structure (Figure S3).

**Phylogenetic tree**

The two cp genome datasets (the 17 complete chloroplast genome sequences assembled in this study and the 99 selected chloroplast genome sequences) were used to reconstruct corresponding phylogenetic trees. As shown in Figure 3, the 17 newly sequenced \textit{Ae. tauschii} accessions clustered into three groups, with Group I being a sister to group II and group III, which formed a single clade. Group I comprised two accessions from Iran and Turkey, which derived from the L2 lineage, whereas the accessions in groups II and III all originated from the L1 lineage. Further, group II included two accessions from Tajikistan, three accessions from Afghanistan, Turkey and Georgia. Group III encompassed two accessions from the Yellow River region (SC1 and T093), three accessions from Xinxiang (XJ0109, XJ04 and XJ098), and a further five accessions from India, Iran, Turkmenistan, Turkistan and Pakistan, respectively.
The phylogenomic tree of Poaceae family, inferred based on the chloroplast genome sequences of 99 selected accessions, was characterized by the two genetic clusters, namely *Triticum* and *Aegilops* clusters (Figure 4). However, we found that *Ae. speltoides* was clustered in a polytomy together with almost all the *Triticum* species, whereas *Triticum urartu* and *Triticum monococcum* grouped with the remaining *Aegilops* species in the *Aegilops* cluster. Furthermore, all D-genome species, including *Ae. tauschii*, *Ae. ventricosa* and *Ae. Cylindrica*, were found to cluster in a single clade.

**Sequence divergences of *Ae. tauschii* chloroplast genomes and molecular marker development**

In order to elucidate variant characteristics among the 17 *Ae. tauschii* chloroplast genomes, we further analyzed divergences in the sequences of coding genes, intron regions and intergenic regions. We accordingly detected 56 variation loci among which 38 loci were located in non-coding regions (34 intergenic regions and four intron regions) and the remaining 18 loci were found in coding regions. A relatively high value of nucleotide variability (\(P_i\)) was thus determined for non-coding regions, ranging from 0.00008 to 0.00635 with an average of 0.00133, which was approximately three times greater than that in the coding regions (average: 0.000432) (Table S3). Among these variable loci, rpl32-trnL-UAG (0.00478), ccsA-ndhD (0.00483), rbcL-psaI (0.00492), and rps18-rpl20 (0.00635), which are located in intergenic regions (the former two in the SSC region and the latter two in the LSC region), displayed the highest nucleotide polymorphisms (Figure 5). Moreover, the primers sequences of four regions can be effectively amplified after verification (Table S4), which will be potential for studying the intraspecific genetic structure and diversity of *Ae. tauschii*.

**Discussion**

In this study, we sequenced the chloroplast genomes of 17 geographically dispersed *Ae. tauschii* accessions, all of which display the typical angiosperms structure and harbor an identical component of 106 unigenes arrayed in the same order (Figure 2). These genomes were found to be relatively conserved, with the IR regions showing greater conservation than either the LSC or SSC region (Figure S1). Differences in genome size could mainly be attributed to intraspecific differences among the cp genomes rather than expansion and contraction.
of IR regions. We found that the size of the 17 assembled genomes tended to be similar, ranging from 135,551 bp to 136,009 bp (Table 2), with accessions AY22 and AY46 having the largest and the smallest genomes, respectively. The IRs regions of all 17 genomes had an identical length (21,548 bp) and differences in genome sequence length can largely be attributed to the variation in the non-coding regions, particularly with respect to the size of the intergenic regions size (Table 2). Comparison with the reference genome indicated that there has been no loss of genes in any of the 17 analyzed genomes. In contrast with findings at the interspecific level statement (Terakami et al., 2012), we detected identical LSC/IR and SSC/IR border regions in the 17 *Ae. tauschii* genomes, thereby indicating that these genomic features would be of little use as evolutionary tools for analyses at the intraspecific level. In conclusion, we found that the general structure of the 17 *Ae. tauschii* cp genomes including gene order and number, has been well conserved. The few variations that we detected are located in the non-coding regions (intergenic spacer regions).

Given the lower nucleotide substitution rates in chloroplast genomes compared with nuclear genomes, it is generally assumed that a substantial sequence length is required for a robust phylogenetic analysis based on chloroplast genome sequences (Wolfe et al., 1987; Khakhlova and Bock, 2006). Accordingly, in order to acquire more comprehensive information, the complete chloroplast genome sequence are beneficial for investigating the phylogenetic relationship of angiosperms (Kim et al., 2015). With respect to the origin of *Ae. tauschii* in China, Yili area of Xinjiang is undisputedly considered to represent the easternmost boundary of the natural distribution of wild *Ae. tauschii* population (Matsuoka et al., 2015; Gogniashvili et al., 2016; Wang et al., 2013), whereas the origin of landraces in the Yellow River region remains a source of debate. In the study, Group I was found derived from the L2 lineage, whereas the accessions in groups II and III all originated from the L1 lineage, which was consistent with the findings of analyses based on 7185 SNP markers (Wang et al., 2013), and thereby indicated that complete chloroplast genome sequences are as equally effective as nuclear sequences for the assessment of the phylogenetic relationships among *Ae. tauschii* accessions. We also found that the Yellow River accessions (SC1 and T093) clustered closely in a small branch with AY22 from Pakistan (South Asia), whereas Xinjiang appeared to clustered together with Central Asian (Figure 3). Thus, the aforementioned results imply that *Ae. tauschii* accessions collected from the Yellow River region (Henan and Shaanxi) could have originated from South Asian populations rather than from those in Xinjiang.
province in China (Wang et al., 2010; Yen et al., 1984) or Iran (Wei et al., 2008), which could be contribute to more effective utilization of wild germplasm resources.

Iran is believed to be the origin of *Ae. tauschii* in the Yellow River region according to the results of Wei et al., (2008). It is conjectured that the spreading route of qingke barley from Western Eurasia though South Asia and into Northern Tibet (Zeng et al., 2018) and the dispersal route of common wheat from the margins of the Qinghai–Tibetan plateau to the Yangtze valley (Wu et al., 2019) may offer valuable clues as to the introduction of *Ae. tauschii* in China (Betts et al., 2014). In addition, some southwest Asian domesticated accessions, including those of pea and rye, appear at Changguogou in the Yarlung Tsangpo river basin of southern Tibet (Fu et al., 2000), and flax is found at Ashaonao on the southeastern Tibetan plateau (Guedes et al., 2014; Guedes et al., 2015), thereby indicating that the introduction of *Ae. tauschii* from South Asia is also possible. Thus, we proposed that the Yellow River accessions may have been introduced directly from South Asia.

Given the aforementioned considerations, we can tentatively propose the possible introduction route whereby *Ae. tauschii* dispersing eastward to China. Specifically, we believe that *Ae. tauschii* accessions distributed in Iran spread eastwards in two routes. One is assumed to have followed a route to Central Asia and Yili of Xinjiang through either human activities or natural extension, whereas the other one migrated eastwards, spreading to South Asia through Northern Tibet, most likely concomitant with the introduction of qingke barley into the Yellow River region, and gradually evolved into *Ae. tauschii* accessions found in this region today. Nevertheless, in order to verify this conjecture, it would be necessary to examine a larger number of *Ae. tauschii* accessions originating from sub-divided areas to facilitate a more comprehensive population analysis, and thereby enable us to trace the exact route of *Ae. tauschii* dispersal into China. In the present study, we investigated genetic variation of *Ae. tauschii* and the source of Chinese landraces, which not only provided a series of new insights to clarify the spreading route of *Ae. tauschii*, but also could be contribute to more effective utilization of wild germplasm resources for the genetic improvement of bread wheat.

In order to gain further insights on the origin of the D genome origin of common wheat, we we performed phylogenetic analysis based on the chloroplast genomes of 99 accession in the *Poaceae* family (Figure 4), and
found that the tree we constructed is similar in topological structure to that presented by Bernhardt (2017).

Most researchers consider that Ae. tauschii, the donator of the D genome of common wheat, could be derived from monophyletic speciation (Dvorak et al., 1998, 2012; Wang et al., 2013; Matsuoka et al., 2015). However, the findings of a study in which the evolutionary relationship of the A, B and D genome lineages were assessed, based on the genome sequences (2269 genes) of hexaploid bread wheat subgenomes and five diploid relatives, indicated that the D genome is derived from homoploid hybrid speciation of the A and B genomes (Marcussen et al., 2014). On the basis of a phylogenetic analysis of chloroplast DNAs, Li et al., (2018) also conjectured that the maternal origin of the D genome lineage might be the A genome or some other relatively close lineage through ancient hybridization. In the present study, with regard to T. urartu (A genome) and Ae. speltoides (B genome), we found that Ae. tauschii firstly clustered together with the other Aegilops specie. If the D genome is derived from an ancient hybridization between the A and B genomes, Ae. tauschii should be closely cluster into T. urartu (the A genome acts a maternal parent) or Ae. speltoides ( the A genome acts a maternal parent).

Thus, we proposed that Ae. tauschii is derived from monophyletic speciation rather than ancient hybridization.

In brief, these phylogeny results will serve as a reference framework for future studies on Triticeae or Ae. tauschii.

To identify the genetic divergence in the assembled genomes, we determined the nucleotide variability (Pi) of coding genes, intron regions and intergenic regions using DnaSP. The results revealed that the sequence divergence of the IR regions appeared to be lower than that of the LSC and SSC regions, which has also been noted in other angiosperms and may possibly be attributed to copy correction of the IR regions via gene conversion (Khakhlova and Bock, 2006). The IR sequences of 17 accessions were identical in this study. Intraspecies variability were detected predominantly in the non-coding regions (intergenic spacer regions) of LSC and SSC (Figure S2 and Figure 5). We also found that four variable loci (rpl32-trnL-UAG, ccsA-ndhD, rbcL-psaI and rps18-rpl20) located in non-coding regions showed notably high levels of nucleotide polymorphisms, two of which (rpl32-trnL-UAG, ccsA-ndhD) are located in the SSC region and the other two loci (rbcL-psaI, rps18-rpl20) are found in the LSC region (Figure 5). Previous studies have identified Ae. tauschii accessions using the chloroplast non-coding sequences trnF-ndhJ, trnC-rpoB, atp1-atpH, and ndhF-rpl32 (Yamane and Kawahara, 2005; Dudnikov, 2012). Here, we developed a further four marker regions (rpl32-trnL-UAG, ccsA-
ndhD, rbcL-psaI and rps18-rpl20) with relatively high levels of intraspecific variation, which can be used for population genetic analyses or serve as specific DNA barcodes of *Ae. tauschii* (Zhang et al., 2011; Maier et al., 1995).

**Conclusion**

In this study, with respect to gene order, gene number, and IR/SC boundary regions, we demonstrate that genomic structure is well conserved among the chloroplast genomes of the 17 *Ae. tauschii* accessions we assembled using Illumina next-generation DNA sequencing technology. Intraspecific differences among the chloroplast genomes were detected primarily in non-coding regions (intergenic spacer regions) which are the main features contributing to the observed differences in genome size. Analysis of the phylogenetic relationships among the accessions based on the complete genome sequences indicated that *Ae. tauschii* accessions from the Yellow River region in China might have originated directly from South Asia. We also confirmed that *Ae. tauschii* is derived from monophyletic speciation rather than hybrid speciation. Furthermore, we identified four cpDNA markers sequences (*rpl32-trnL-UAG*, *ccsA-ndhD*, *rbcL-psaI*, and *rps18-rpl20*) that can be used to study inter- and intraspecific genetic structure and diversity of *Ae. tauschii*.

**Additional Information And Declarations**

**Funding**

This work was supported by National Natural Science Foundation of China (Grant Nos. 31601297 and 31871615) and Key Scientific Research Projects of Higher Education Institutions in Henan Province (Grant No. 20A210007).

**Competing interests**

The authors declare no potential conflict of interest.

**DNA Deposition**
The following information was supplied regarding the deposition of DNA sequences: These chloroplast genome sequences have been successfully submitted into GenBank with the accession numbers in the following: MN258085 (SC1); MN258083 (AY81); MN258078 (AY34); MN223978 (AY22); MN258084 (AY320); MN223977 (AY21); MN258087 (XJ04); MN258089 (XJ0109); MN258086 (T093);  MN258088 (XJ098); MN258082 (AY78); MN223975 (AS060); MN258080 (AY48); MN258090 (AY076); MN223976 (AY20); MN258081 (AY72); MN258079 (AY46).

Data Availability

The following information was supplied regarding data availability:

The raw data are provided in the Supplemental Files.

References

Bernhardt N, Brassac J, Kilian B, and Blattner FR. 2017. Dated tribe-wide whole chloroplast genome phylogeny indicates recurrent hybridizations within Triticeae. *BMC Evolutionary Biology* 17(1):141.

Betts A, Jia PW and Dodson J. 2014. The origins of wheat and potential pathways for its introduction: A review. *Quaternary International* 348: 158-168.

Dodson JR, Li X, Zhou X, Zhao K, and Atahan P. 2013. Origin and spread of wheat in china. *Quaternary Science Reviews* 72(2):108-111.

Dudnikov AJ. 2012. Chloroplast DNA non-coding sequences variation in *Aegilops tauschii* Coss.: evolutionary history of the species. *Genetic Resources and Crop Evolution* 59:683-699.

Dvorak J, Deal KR, Luo MC, You FM, Borstel KV, and Dehghani H. 2012. The origin of spelt and free-threshing hexaploid wheat. *Journal of Heredity* 103:426-441.

Dvorak J, Luo MC, Yang ZL, and Zhang HB. 1998. The structure of the Ae. *tauschii* genepool and the evolution of hexaploid wheat. *Theoretical and Applied Genetics* 97:657-670.

Fu DX, Xu TW, and Feng ZY. 2000. The ancient carbonized barley (*Hordeum vulgare* L. Var. nudum) kernel discovered in the middle of Yalu Tsanypo river basin in Tibet. *Southwest China Journal of Agriculturalences* 13: 38-41.

George B, Bhatt BS, Awasthi M, George B, and Singh AK. 2015. Comparative analysis of microsatellites in
chloroplast genomes of lower and higher plants. *Current Genetics* 61:665-677.

Gogniashvili M, Jinjikhadze T, Maisaia I, Akhalkatsi M, Kotorashvili A, and Kotaria N. 2016. Complete chloroplast genomes of *Aegilops tauschii* Coss. and *Aegilops cylindrica* host sheds light on plasmon D evolution. *Current Genetics* 62:791-798.

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

Guedes JA, Lu HL, Hein AM, and Schmidt AH. 2015. Early evidence for the use of wheat and barley as staple crops on the margins of the tibetan plateau. *Proceedings of the National Academy of Sciences* 112(18): 5625-5630.

Guedes JA, Lu HL, Li YX, Spengler RN, Wu XH, and Aldenderfer MS. 2014. Moving agriculture onto the Tibetan plateau: the archaeobotanical evidence. *Archaeological and Anthropological Sciences* 6: 255-269.

Han Y, Chen W, Yang F, Wang X, Dong M, Zhou P, and Shang F. 2015. cDNA-AFLP analysis on *Osmanthus fragrans* cultivars with different flower color and molecular characteristics of OfMYB1 gene. *Trees* 29(3): 931-940.

Jaaska V. 1980. Electrophoretic survey of seedling esterases in wheats in relation to their phylogeny. *Theoretical and Applied Genetics* 56:273-284.

Katoh K, Misawa K, Kuma KI, and Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic acids research* 30:3059-3066.

Khakhlova O, and Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant Journal* 46:85-94.

Kihara H, Yamashita H, and Tanaka M. 1965. Morhologic, physiological, genetical, and cytological studies in *Ae.* and *Triticum* collected in Pakistan, Afghanistan, Iran; Kyoto University Press: Kyoto, 4-41.

Kilian B, Mammen K, Millet E, Sharma R, Graner A, Salamini F, Hammer K, and Ozkan H. 2011. Wild crop relatives: genomic and breeding resources: Cereals. *Springer* 1-76.

Kim KJ, and Lee HL. 2004. Complete chloroplast genome sequences from Korean ginseng (*Panaxschinseng Nees*) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Research* 2004 11:247-261.

Kim K, Lee SC, Lee J, Yu Y, Yang K, and Choi BS. 2015. Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Scientific Reports* 5:15655.

Kimura M. 1984. The neutral theory of molecular evolution. *Current Opinion in Genetics and Development*
Kumar S, Stecher G, and Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 33:1870-1874.

Lascano HR, Casano LM, Martin M, and Sabater B. 2003. The activity of the chloroplast Ndh complex is regulated by phosphorylation of the NDH-F subunit. *Plant Physiology* 13: 2256-2262.

Li CP, Sun XH, Conover JL., Zhang ZB, Wang JB, Wang XF, Deng X, Wang HY, Liu B, Wendel JF and Gong L. 2018. Cytonuclear Coevolution following Homoploid Hybrid Speciation in *Ae. tauschii* [J]. *Molecular biology and evolution*. Doi:10.1093/molbev/msy215.

Li JJ, Liu HX, Mao SZ, Zhao B, and Huang SX. 2016. Adaptive evolution of the *ndhF* gene in the genus *Rheum* (Polygonaceae). *Guihaia* 36:101-106.

Librado P, and Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.

Liu LX, Li R, Worth JR, Li X, Li P, Cameron KM, and Fu CX. 2017. The complete chloroplast genome of Chinese Bayberry (*Morella rubra*, Myricaceae): implications for understanding the evolution of Fagales. *Frontiers in Plant Science* 8:968.

Liu LX, Li P, Zhang HW, Worth JRP. 2018. Whole chloroplast genome sequences of the Japanese hemlocks, *Tsuga diversifolia* and *T-sieboldii*, and development of chloroplast microsatellite markers applicable to East Asian *Tsuga*. *Journal of Forest Research* 23(5): 318-323.

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:W575.

Long T, Leipe C, Jin GY, Wagner M, Guo RZ, Schröder O, and Tarasov PE. 2018. The early history of wheat in China from 14C dating and Bayesian chronological modelling. *Nature Plants* 4: 272-279.

Lubbers EL, Gill KS, Cox TS, and Gill BS. 1991. Variation of molecular markers among geographically diverse accessions of *Triticum tauschii*. *Genome* 34:354-361.

Maier RM, Neckermann K, Igloi GL, 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. *International Journal of Biological Macromolecules* 251: 614-628.

Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, Jakobsen K S, Wulff BBH, Steuernagel B, Msyer KFX, and Olsen O-A. 2014. Ancient hybridizations among the ancestral genomes of bread wheat. *Science* 345(6194):1250092.

Matsuoka Y, Mori N, and Kawahara T. 2005. Genealogical use of chloroplast DNA variation for intraspecific studies of *Ae. tauschii* Coss. *Theoretical and Applied Genetics* 111:265-271.
Matsuoka Y, Takumi S, and Kawahara T. 2015. Intraspecific lineage divergence and its association with reproductive trait change during species range expansion in central Eurasian wild wheat *Aegilops tauschii* Coss. (Poaceae). *BMC Evolutionary Biology* 15:1-10.

Middleton CP, Senerchia N, Stein N, Akhunov ED, Keller B, Wicker T, and Benjamin K. 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:e85761.

Mizuno N, Yamasaki M, Matsuoka Y, Kawahara T, and Takumi S. 2010. Population structure of wild wheat D-genome progenitor *Aegilops tauschii* Coss.: implications for intraspecific lineage diversification and evolution of common wheat. *Molecular Ecology* 19:999-1013.

Palmer JD. 1991. Plastid chromosomes: structure and evolution. *The molecular biology of plastids* 7: 5-53.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Molecular biology and evolution* 25:1253-1256.

Sang T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* 37:121-147.

Shang, Zang BS, Zhao, YH, Zhu X, Fan, JD, Guo C, Zhang XN, Xiao. 2019. Functional characterization of GhPHOT2 in chloroplast avoidance of *Gossypium hirsutum*. Plant Physiol Biochem 135: 51-60.

Singh S, Chahal GS, Singh PK, and Gill BS. 2012. Discovery of desirable genes in the germplasm pool of *Ae. tauschii* Coss. *Indian Journal of Genetics and Plant Breeding* 72:271-277.

Tabidze V, Baramidze G, Pipia I, Gogniashvili M, Ujmajuridze L, Beridze T, Hernandez AG, and Schaal B. 2014. The complete chloroplast DNA sequence of eleven grape cultivars. Simultaneous resequencing methodology. *Journal International Des Sciences De La Vigne Et Du Vin* 48: 99-109.

Terakami S, Matsumura Y, Kurita K, Kanamori H, Katayose Y, Yamamoto T, and Katayama H. 2012. Complete sequence of the chloroplast genome from pear (*Pyrus pyrifolia*): genome structure and comparative analysis. *Tree Genetics and Genomes* 8(4):841-854.

Wang J, Luo MC, Chen Z, You FM, Wei Y, and Zheng Y. 2013. *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. *New Phytologist* 198:925-937.

Wang Q, Huang L, Yuan ZW, Hu XG, and Liu DC. 2010. Transmission relationship between *Aegilops tauschii* Cossions in Xinjiang and Yellow River Basin in China. *Journal of Sichuan Agricultural University* 28: 407-410.

Wei HT, Li J, Peng ZS, Lu BR, Zhao ZJ, and Yang WY. 2008. Relationships of *Aegilops tauschii* revealed by DNA fingerprints: The evidence for agriculture exchange between China and the West. *Progress in Natural Science-materials International* 18:1525-1531.

Wolfe KH, Li WH, and Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant
mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences* 84:9054-9058.

Wu XY, Ding BX, Zhang BQ, Feng JJ, Wang YB, Ning C, Wu HD, Zhang F, Zhang Q, Li N, Zhang ZB, Sun XH, Zhang QC, Li WY, Liu B, Cui YQ, and Gong L. 2019. Phylogenetic and population structural inference from genomic ancestry maintained in presen-day common wheat Chinese Landraces. *The Plant Journal* 99(2): doi: 10.1111/tpj.14421.

Wyman SK, Jansen RK, and Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252-3255.

Yamane K, and Kawahara T. 2005. Intra- and interspecific phylogenetic relationships among diploid *Triticum-Aegilops* species (Poaceae) based on base-pair substitutions, indels, and microsatellites in chloroplast non-coding sequences. *American Journal of Botany* 92:1887-1898.

Yang M, Zhang X, Liu G, Yin YX, Chen KF, Yun QZ, Zhao DJ, and Yu J. 2010. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). *PLoS ONE* 5:e12762.

Yen C, Cui NR, and Dong YS. 1984. The *Aegilops tauschii* coossion from yi-li, xinjiang, china. *Acta Agronomica Sinica* 1:1-8.

Yen C, Yang JL, Liu XD, and Li LR. 1983. The distribution of *Aegilops tauschii* Cosson in China and with reference to the origin of the Chinese common wheat. *International Wheat Genetics Symposium* 55-58.

Zeng XQ, Guo Y, Xu QJ, Mascher M, Guo GG, Li SC, Mao LK, Liu QF, Xia ZF, Zhou JH, Yuan HJ, Tai SS, Wang YL, Wei ZX, Song L, Zha S, Li SM, Tang YW, Bai LJ, Zhuang ZH, He WM, Zhao SC, Fang XD, Gao Q, Yin Y, Wang J, Yang HM, Zhang J, Henry RJ, Stein N, and Tashi N. 2018. Origin and evolution of qingke barley in Tibet. *Nature Communications* 9: https://doi.org/10.1038/s41467-018-07920-5.

Zhang DL, He J, Zhang CC, Zhou Y, Su YR, and Li SP. 2017. An advanced backcross population through synthetic octaploid wheat as a “bridge”: development and QTL detection for seed dormancy. *Frontier in Plant Science* 8:2123.

Zhang DL, Zhou Y, Zhao XP, Lv LL, Zhang CC, Li JH, Sun GL, Li SP, and Song CP. 2018. Development and Utilization of Introgression Lines Using Synthetic Octaploid Wheat (*Aegilops tauschii* × Hexaploid Wheat) as Donor. *Frontier in Plant Science* 9: 1113.

Zhang YJ, Ma PF, and Li DZ. 2011. High-throughput sequencing of six bamboo chloroplast genomes: Phylogenetic implications for temperate woody bamboos (*Poaceae: Bambusoideae*). *PLoS ONE* 6:e20596.
Figure 1

Geographical distribution of 17 *Aegilops tauschii* accessions from western Turkey to eastern China
Figure 2

Chloroplast genome map of 17 *Aegilops tauschii* accessions.

The genes lying outside and inside of circle are transcribed in the counterclockwise and clockwise directions, respectively. Different colored bars are used to represent different functional gene groups. The darker gray area in the inner circle displays GC content. The fine lines indicate the boundary of the inverted repeats (IR$_A$ and IR$_B$) that split the genomes into large single copy (LSC) and small single copy (SSC) regions.
Figure 3

Phylogenetic tree constructed based on the complete chloroplast genomes of 17 *Aegililops tauschii* accessions by maximum likehood (ML) method.

Bootstrap support values (<50%) were hided. The phylogenetic tree resulting from analysis of 135 984 bp in the alignment length of chloroplast genomes with all gap positions removed, including long stretches of the same nucleotide, short sequences appearing in opposite orientation and some sequences consisting short repeats.
Figure 4

Phylogenetic tree constructed based on the complete chloroplast genomes of 99 Triticeae accessions by maximum likehood (ML) method.

The phylogenetic tree resulting from analysis of 131,116 bp in the alignment length of chloroplast genomes with all gap positions removed, including long stretches of the same nucleotide, short sequences appearing in opposite orientation and some sequences consisting short repeats.
Manuscript to be reviewed
Figure 5

Comparative analysis of the nucleotide variability (Pi) values among 17 *Aegilops tauschii* accessions.

The homologous regions are oriented according to their locations in the chloroplast genomes.
| Accession | Origin | Collection Region |
|-----------|--------|-------------------|
| Aegilops tauschii | Asia | China |
| | Europe | Greece |
| | | Italy |

Note: Table 1 continues on the next page.
### Table 1

| Accessions* | Inventory** | Source | Regions |
|-------------|-------------|--------|---------|
| SC1         |             | Shannxi, China | N (34.158997), E (108.90699), Elevation: 428 meters |
| AY81        | PI 542277   | Izmir, Turkey | Elevation: 30 meters |
| AY34        | PI 662116   | Khujand, Tajikistan | N (39.771944), E (68.809444), Elevation: 433 meters |
| AY22        | PI 511365   | Baluchistan, Pakistan | N (30.925), E (66.44638889), Elevation: 675 meters |
|             |             |               | N (40.13333333), E (43.06666667), Elevation: 1275 |
| AY320       | PI 554324   | Kars, Turkey | meters |
| AY21        | PI 662091   | Khujand, Tajikistan | N (40.67388889), E (70.54694444), Elevation: 462 meters |
| XJ04        |             | Xinjiang, China | N (44.321239), E (80.77766), Elevation: 892 meters |
| XJ0109      |             | Xinjiang, China | N (43.386026), E (83.5977), Elevation: 1269 meters |
| T093        |             | Henan, China | N (35.728123), E (115.242698), Elevation: 52 meters |
| XJ098       |             | Xinjiang, China | N (43.386026), E (83.5977), Elevation: 1269 meters |
| AY78        | PI 210987   | Kondoz, Afghanistan | N (36.68333333), E (68.05), Elevation: 362 meters |
| AS060       | PI 511369   | Iran | N(36.85), E(55.170593), Elevation: 527 meters |
| AY48        | PI 603225   | Turkmenistan, Balkan | N(38.48333333), E(56.3), Elevation: 730 meters |
| AY076       | PI 276975   | Turkistan | N (45), E (70), Elevation: 210 meters |
| AY20        | PI 574469   | India | N(20), E(77), Elevation: 509 meters |
AY72  PI 428563  Georgia  N(43),E(47), Elevation: 90 meters

AY46  PI 511368  Tehran, Iran  N(35.8), E(50.9666667), Elevation: 1296 meters

Note: * respect the accession numbers from Key Laboratory of Plant Stress Biology of Henan University.

** respect the inventory provided by the US National Plant Germplasm Center and Institute of Genetics and Developmental Biology.
Table 2 (on next page)

Comparative analysis of the chloroplast genomes among 17 *Aegilops tauschii* accessions
| Accessions | AY21 | AY22 | AY320 | AY34 | AY81 | SC1 | T093 | XJ04 | XJ098 | XJ0109 | XJ078 | AS060 | AY20 | AY72 | AY48 | AY076 | AY46 |
|------------|------|------|-------|------|------|-----|------|------|-------|--------|-------|-------|------|------|------|------|------|
| Total size(bp) | 135858 | 136009 | 135777 | 135608 | 135890 | 135656 | 135617 | 135813 | 135836 | 135854 | 135551 | 135413 | 135412 | 135411 | 135410 | 135409 |
| GC% | 38.32 | 38.11 | 38.32 | 38.32 | 38.32 | 38.32 | 38.32 | 38.33 | 38.33 | 38.33 | 38.33 | 38.33 | 38.33 | 38.33 | 38.33 | 38.33 |
| Gene total length(bp) | 59706 | 59703 | 59628 | 59703 | 59703 | 59703 | 59703 | 59703 | 59703 | 59703 | 59703 | 59703 | 59703 | 59703 | 59703 | 59703 |
| Gene average length(bp) | 319 | 319 | 319 | 319 | 319 | 319 | 319 | 319 | 319 | 319 | 319 | 319 | 319 | 319 | 319 | 319 |
| Gene density (genes per kb) | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 | 0.61 |
| GC content in gene region(%) | 38.9 | 38.9 | 38.9 | 38.9 | 38.9 | 38.9 | 38.9 | 38.9 | 38.9 | 38.9 | 38.9 | 38.9 | 38.9 | 38.9 | 38.9 | 38.9 |
| Gene/Genome(%) | 43.9 | 43.9 | 43.8 | 44.0 | 43.9 | 44.0 | 43.9 | 44.0 | 43.9 | 44.0 | 43.9 | 44.0 | 43.9 | 44.0 | 43.9 | 44.0 |
| Intergenic region length (bp) | 76152 | 76306 | 76419 | 76074 | 76262 | 75905 | 72275 | 77027 | 75908 | 75908 | 75908 | 75908 | 75908 | 75908 | 75908 | 75908 |
| GC content in intergenic region(%) | 37.8 | 37.8 | 37.8 | 37.8 | 37.8 | 37.8 | 37.8 | 37.8 | 37.8 | 37.8 | 37.8 | 37.8 | 37.8 | 37.8 | 37.8 | 37.8 |
| Intergenic length/Genome(%) | 56.1 | 56.1 | 56.1 | 56.1 | 56.1 | 56.1 | 56.1 | 56.1 | 56.1 | 56.1 | 56.1 | 56.1 | 56.1 | 56.1 | 56.1 | 56.1 |
| LSC(bp) | 79991 | 80142 | 80111 | 79910 | 80020 | 79741 | 79983 | 79744 | 79743 | 79746 | 79789 | 79749 | 79946 | 79969 | 79987 | 79723 |
| SSC(bp) | 12771 | 12771 | 12771 | 12771 | 12771 | 12771 | 12771 | 12771 | 12771 | 12771 | 12771 | 12771 | 12771 | 12771 | 12771 | 12771 |
| IR(bp) | 21548 | 21548 | 21548 | 21548 | 21548 | 21548 | 21548 | 21548 | 21548 | 21548 | 21548 | 21548 | 21548 | 21548 | 21548 | 21548 |
| Total number of genes | 122 | 122 | 122 | 122 | 122 | 122 | 122 | 122 | 122 | 122 | 122 | 122 | 122 | 122 | 122 | 122 |
| Number of protein-coding genes | 82(6) | 82(6) | 82(6) | 82(6) | 82(6) | 82(6) | 82(6) | 82(6) | 82(6) | 82(6) | 82(6) | 82(6) | 82(6) | 82(6) | 82(6) | 82(6) |
| Number of tRNA genes | 32(6) | 32(6) | 32(6) | 32(6) | 32(6) | 32(6) | 32(6) | 32(6) | 32(6) | 32(6) | 32(6) | 32(6) | 32(6) | 32(6) | 32(6) | 32(6) |
Duplicated genes in IR

2

3

4