Genet Resour Crop Evol
https://doi.org/10.1007/s10722-023-01843-x

RESEARCH ARTICLE

Sequencing and analysis of complete chloroplast genomes of einkorn wheats *Triticum sinskajae* and *Triticum monococcum* accession k-20970

Kuluev Azat · Kuluev Bulat · Mikhaylova Elena · Chemeris Alexey

Received: 17 November 2023 / Accepted: 18 December 2023

Abstract *Triticum sinskajae* A. Filat. et Kurk. has been discovered in the early 1970s during the reproduction of *Triticum monococcum* L. accession k-20970. It is still debated whether it is a new species or a spontaneous mutant of *T. monococcum*. Although *T. sinskajae* appears to be genetically distinct from several *T. monococcum* lineages, there is still no data on the differences from accession k-20970, from which it may have originated. The purpose of this study was to compare the chloroplast genomes of *T. sinskajae* and *T. monococcum* accession k-20970 and clarify the phylogenetic relationships between einkorn wheats. Complete chloroplast genomes of *T. sinskajae* k-48993 (Dagestan), *T. monococcum* k-20970 (Turkey), *T. boeoticum* k-58674 (Armenia) have been sequenced and annotated for the first time. Chloroplast genome sequences of *T. monococcum* subsp. monococcum and *T. urartu* were used for comparative analysis. However *T. monococcum* k-20970 and *T. sinskajae* k-48993 had common mutations that were absent in *T. monococcum* subsp. monococcum, they also were polymorphic in 10 nucleotides, including 1 transversion, 1 deletion, 4 insertions and 4 bp inversion of AGAA to TTCT in the *rbcL-psaI* intergenic region. Due to genetic and phenotypic differences *T. sinskajae* k-48993 and *T. monococcum* k-20970 can be considered different species. A comparison of the chloroplast genomes of einkorn wheats showed a common origin of the plastomes of three closely related species —*T. sinskajae*, *T. monococcum* and *T. boeoticum*, while *T. urartu* can be assigned to another clade.

Keywords *Triticum sinskajae* · Einkorn wheat · Chloroplast genome · Phylogeny · Triticum-Aegilops alliance

Introduction

Einkorn wheat is a group of three diploid species with genome A: *Triticum monococcum* L., *Triticum boeoticum* Boiss. and *Triticum urartu* Thum. ex Gandil. *Triticum sinskajae* A. Filat. et Kurk. (Sinskaya wheat) is sometimes considered the fourth species of this group. It has been discovered in the early 1970s during the reproduction of *T. monococcum* samples (accession k-20970) at the Central Asian and Dagestan VIR stations. Samples of accession k-20970 were brought by P.M. Zhukovsky from surroundings of the city of Daday, Kastamonu province (Turkey) in 1926 (Filatenko and Kurkiev 1975).

Sinskaya wheat grows up to 120 cm in height. *T. sinskajae* differs from *T. monococcum* morphologically, by soft ear glumes, shorter but denser ears, larger spikelets, and a less developed awn (Filatenko and Kurkiev 1975). Glossy smooth glumes of

K. Azat (✉) · K. Bulat · M. Elena · C. Alexey
Institute of Biochemistry and Genetics, Ufa Federal Research Centre of RAS, Ufa, Russia
e-mail: kuluev.azat91@yandex.ru

Sinskaya wheat are longer and wider than those of *T. monococcum*, and it has compact and non-fragile spike (Watanabe 2017; Vavilova et al. 2020). *T. sinskajae* is the only free-threshing species among einkorn wheats (Dorofeev et al. 1979; Simons et al. 2006). There is a suggestion that Sinskaya wheat could have occurred as a result of a spontaneous mutation of *T. monococcum* (Kurkiev and Filatenko 2000). It turned out that in *T. sinskajae* there is an additional very narrow scale between the ear glume and outer flower scales. Authors described that on the ear there is a small step and suggested that *T. sinskajae* originated as a mutant of *T. monococcum* in which functions of poorly developed soft ear glumes were performed by the lower floral glumes. According to Kurkiev and Filatenko (2000), this mutation could occur through the loss of a chromosome locus with a block of genes.

Since *T. sinskajae* is a poorly studied plant, rarely included in phylogenetic studies of the *Triticeae* tribe, it is important to sequence its nuclear and chloroplast genomes. It is also of great interest to study its phylogenetic position in the Triticum-Aegilops alliance. The chloroplast genome of another einkorn wheat *T. boeoticum* also has not been completely sequenced. Previously, sequences of a number of marker genes were compared in *T. monococcum* accession k-20970 and *T. sinskajae*. Analysis of the nuclear gene *Acc-1* (acetyl-CoA carboxylase) revealed a 46 bp deletion in intron 11 of *T. monococcum* (k-20970), which was absent in in *T. sinskajae* (k-48993). Moreover, *T. monococcum* (k-20970) did not have a T/C substitution at position 3709, which was common for *T. sinskajae* (k-48993) and several other *T. monococcum* samples. Based on these results, the authors of the study suggested that there is not yet enough evidence to divide *T. monococcum* and *T. sinskajae* into two distinct species (Golovnina et al. 2009).

We have previously sequenced the chloroplast genome of *T. sinskajae*. Comparison of the chloroplast genomes of *T. sinskajae* and *T. monococcum* subsp. monococcum (LC005977) showed 99.96% similarity between the samples. The total number of polymorphisms was 48, which supports the recognition of *T. sinskajae* as species (Kuluev et al. 2020). However, plastome of its possible ancestor *T. monococcum* accession k-20970 has not yet been sequenced. Based on the above, the objective of our research was to sequence and compare the chloroplast genomes of *T. sinskajae* and *T. monococcum* accession k-20970, as well as the chloroplast genome of *T. boeoticum* k-58674 (Armenia) in order to clarify the phylogenetic relationships in tribe *Triticae*. Comparison of the chloroplast genomes of *T. sinskajae* and other einkorn wheats and *Aegilops* species may help answer a number of questions concerning the study of phylogenetic relationships in the Triticum-Aegilops alliance.

Materials and methods

Plant material and chloroplast DNA extraction and sequencing

Chloroplast DNA was isolated from 15 g of fresh green leaves of *T. sinskajae*, *T. monococcum* and *T. boeoticum* using modified protocol (Farrar and Donnison 2007). Young green leaves were homogenized with liquid nitrogen in a mortar and transferred into 200 ml of ice-cold HBS buffer: 1xHB stock with addition of 0.5 M sucrose (pH 9.4–9.5). Composition of HB stock: 0.1 M Tris, 0.8 M KCl, 0.1 M EDTA, 10 mM spermidine-HCl, 10 mM spermine-HCl. β-mercaptoethanol was added to the buffer immediately before use. The solution was stirred on ice for 30 min, filtered through two layers of miracloth material consisting of viscose-polyester with an acrylic binder and centrifuged at 4 °C for 2 min at 2000 rpm on a 5804R centrifuge (Eppendorf, Germany). The precipitate was dissolved in 25 ml of HB buffer to reduce the concentration of sucrose. The resulting solution was centrifuged for 15 min at 5000 g. The supernatant was partially poured out and the precipitate was carefully resuspended in the remaining volume of HBS buffer (5–8 ml).

Chloroplasts isolated with a HBS buffer were transferred to ultracentrifuge tubes with sucrose gradient consisting of 30% sucrose carefully loaded onto 50% sucrose. The chloroplast suspension was carefully layered onto the gradient to avoid mixing. Tubes were centrifuged for 45 min at 4 °C and 10,000 g on an Optima L-90 K ultracentrifuge (Beckman Coulter, USA). Intact chloroplasts accumulated between 50 and 30% sucrose layers were collected into test tubes using a Varioperpex II Pump (LKB, Sweden). The resulting solutions were concentrated by centrifugation, resuspended in Triton X-100 (0.15% v/v) and then concentrated again. Chloroplast DNA was isolated from precipitates using the standard method of phenol–chloroform extraction (Graham 1978). The quality of the isolated chloroplast DNA was evaluated using a Nanodrop spectrophotometer (Thermo Fisher Scientific, USA). Ultrasonic chloroplast DNA fragmentation was performed on an M220 focused-ultrasonicator (Covaris, USA). Chloroplast DNA libraries were prepared with the KAPA HyperPrep kit (Roche, Switzerland) and dual-indexed adapters according to the manufacturer's protocol. Paired-end sequencing was performed on a Genolab M sequencer (GeneMind Biosciences, China) using Sequencing Reagent Kit v2.5 (FCM 300 Cycles). A read length of 150 bp and a coverage of 8–10 million reads per sample were achieved. Raw data was filtered via Trimmomatic v0.22 to obtain clean reads (Bolger et al. 2014).

Gene assembly and annotation

The chloroplast genomes were assembled by NOVOwrap (Wu et al. 2021) using the chloroplast genome of *T. monococcum* subsp. *monococcum* (LC005977.1) as a reference. The complete chloroplast genome sequences were annotated using Chloroplast Genome Annotation, Visualization, Analysis, and GenBank Submission 2 web server (CPGAVAS2) (http://47.96.249.172:16019/analyzer/home (accessed on 15 September 2023)) (Shi et al. 2019). The circular genome map was generated using Chloroplot tool (https://irscope.shinyapps.io/Chloroplot/ (accessed on 15 September 2023)) (Zheng et al. 2020). The annotated genes were checked, and the errors were corrected manually. Accession numbers of the chloroplast genomes of analyzed *Triticum* and *Aegilops* species are given in Table 1. Multiple sequence alignments of full chloroplast genomes were performed using MAFFT v7.427 (scoring matrix = 200, PAM = 2, gap open penalty = 1.53, offset value = 0.123, FFT-NS-1 method) (Katoh and Standley 2013). Subsequently, the aligned sequences, boundaries of four regions (large single-copy (LSC), small single-copy (SSC), and a pair of invert repeats (IRs)) of the chloroplast genomes were analyzed by Ugene v48.1 software. Multiple sequence alignments were visualized using Jalview v2.11.2.7 (Waterhouse et al. 2009). Nucleotide numbering was designated according to the consensus sequence generated by the Jalview v2.11.2.7 program.

Phylogenetic analysis

A phylogenetic tree was reconstructed by Archaeopteryx JavaPlugin with 1000 bootstrap repetitions (Han and Zmasek 2009). The *Secale cereale* subsp. segetale chloroplast genome from GenBank (MZ507427) have been chosen as an outgroup species.

Table 1 Some species of wheat and *Aegilops* subjected to comparative analysis of chloroplast genomes

Species and genomes	Accession number
Genus Aegilops, section Sitopsis, subsection Emarginata	NC_024830
Aegilops longissima S^l	NC_024815
Aegilops searsii S^s	NC_024831
Aegilops bicornis S^b	NC_024816
Aegilops sharonensis S^{sh}	KJ614406
subsection Truncata	KJ614405
Aegilops speltoides var. speltoides S^{sp}	KJ614412
Aegilops speltoides var. ligustica S^{sp}	
Section Vertebrata	
Aegilops tauschii D	
Genus *Triticum*	KJ614411
Triticum urartu A^u	LC00597
Triticum monococcum subsp. monococcum A^m	
Triticum turgidum subsp. durum BA	KM352501
Triticum timopheevii GA	KJ614410
Triticum timopheevii subsp. araraticum	LC655300
Triticum aestivum BAD	CM022232
Triticum zhukovskyi GAA	NC_046698

Results

Annotation of the chloroplast genomes of *T. sinskajae, T. monococcum* k-20970 and *T. boeoticum* k-58674

Analysis with CPGAVAS2 revealed that the size of the chloroplast genome of *T. sinskajae* was 136,885 bp, which corresponds to the size of the chloroplast genomes of other cereal species (Zhang et al. 2011; Bernhardt et al. 2017; Gogniashvili et al. 2018; Su et al. 2020). Chloroplast genome of *T. sinskajae* has the typical quadripartite structure with a pair of inverted repeats (IRA- and IRB regions, each 21,547 bp in size), one SSC region (12,809 bp) and one LSC (80,982 bp). The total coding sequence is 73,226 bp. The GC content in the complete chloroplast genome of *T. sinskajae* is 38.3% (39.6% in the SSC, 36.2% in the LSC, 43.9% in the IR). The higher GC content in the IR region can be explained by the presence of four rRNA genes, which is consistent with previous analyses of chloroplast DNA of other plants (Bosacchi et al. 2015; Nie et al. 2018).

In Sinskaya wheat, 130 structural chloroplast genes were annotated: 83 protein-coding genes, 22 tRNA genes, 4 rRNA genes and 1 gene of unknown function (*ycf4*). Among them, 5 protein-coding genes (*rpl2, rpl23, ndhB, rps7, rps12B*), 8 tRNA genes (*trnH-GUG, trnl-CAU, trnL-CAA, trnV-GAC, trnI-GAU, trnA-UGC, trnR-ACG, trnV-GUU*), 4 rRNA genes (*rRNA4.5, rRNA23, rRNA16* and *rRNA5*) were duplicated due to being in the IR repeat region. Besides, 10 protein-coding and 1 tRNA genes were found in SSC, and 62 protein-coding genes and 21 tRNA genes were found in LSC. Out of 130 genes, 11 have one intron (*atpF, ndhB, petB, petD, rpl2, trnI-GAU, ndhA, rpl16, rps12A, rps16, trnG-UCC*) and 9 genes have 2 introns (*trnK-UUU, trnS-CGA, trnL-UAA, trnV-UAC, trnT-CGU, trnA-UGC, ycf3*). Among 17 genes with introns, 2 protein-coding genes and 2 tRNA genes are located in the IR. The SSC region contains the *ndhA* gene with an intron. The LSC region contains 12 genes with introns, including 8 protein-coding genes, 4 tRNA-coding genes, and one gene with two introns (*ycf3*). The largest intron (2479 bp) is located in the *trnK-UUU* gene which includes *matK* gene. All obtained results are presented in Fig. 1 as a circular structure built via the Chloroplot tool. Various color blocks on the outer circle reflect the genes belonging to certain functional groups.

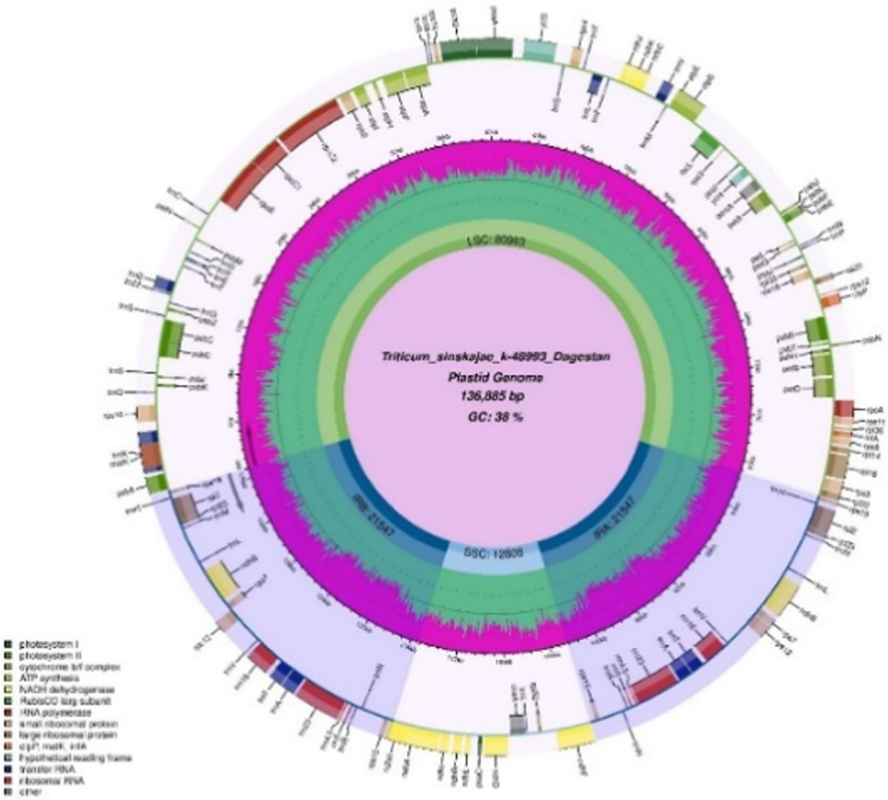
The chloroplast genomes of *T. monococcum* k-20970 and *T. boeoticum* k-58674 were also annotated using CPGAVAS2. The same genes with the same copy number and intron location were found in their chloroplast genomes. There was only one exception: in *T. monococcum* k-20970 and *T. boeoticum* k-58674 there was a *ycf15* gene of unknown function duplicated in the IR region (Table 2). The annotation of the chloroplast genomes of *T. monococcum* k-20970 and *T. boeoticum* k-58674 was visualized using the Chloroplot tool (Figs. 2 and 3).

Basic information on the annotation results of all three analyzed chloroplast genomes of einkorn wheats, as well as data on *T. monococcum* subsp. monococcum (LC005977) and *T. urartu* (KJ614411) are presented in Table 3. Overall, there were few significant differences among all einkorn wheats. *T. boeoticum* had the largest chloroplast genome, and the genome of *T. urartu* was the smallest. *T. monococcum* k-20970 and *T. boeoticum* k-58674 had 2 more genes (2×*ycf15*) in the chloroplast genome than the other three einkorn wheats (Table 3).

Identification of *T. sinskajae*-specific mutations in the chloroplast genome

To identify *T. sinskajae*-specific mutations, pairwise alignment and comparison of nucleotide sequences of einkorn chloroplast genomes of all species of einkorn wheats were performed. The comparison of the chloroplast genomes of *T. sinskajae* and *T. monococcum* subsp. monococcum (LC005977) via BLAST tool showed 99.96% similarity between these species. The total number of polymorphisms was 48. Comparison of the chloroplast genome sequences of *T. sinskajae* and *T. urartu* showed 99.82% similarity with a greater number of polymorphisms (245). Comparison of the chloroplast genome sequences of *T. monococcum* (k-20970) and *T. monococcum* subsp. monococcum, as well as *T. boeoticum* (k-58674) and *T. monococcum* subsp. monococcum using BLAST showed 99.96% similarity in the chloroplast genomes of these species. Alignment of chloroplast genomes of *T. sinskajae* and *T. monococcum* subsp. monococcum using the MAFFT v7.427 program revealed 17 single-nucleotide substitutions, including 5 transitions

Fig. 1 Visual circular representation of the sequenced chloroplast genome of *T. sinskajae* generated using the Chloroplot resource. Genes are shown in different colors; the purple area in the middle of the circle shows the level of GC content. IRA—inverted repeat region (**A**), IRB—inverted repeat region (**B**). Genes located outside the outer circle are transcribed clockwise, and those located inside are transcribed counterclockwise



and 12 transversions, and GAA → TTC inversion at position 107,906–107908. There were 9 single-nucleotide and 2 dinucleotide insertions, 7 single-nucleotide deletions and two deletions ATAGC and TTTT in *T. sinskajae* (20 indels in general), compared with *T. monococcum* subsp. monococcum. Between *T. monococcum* k-20970 and *T. monococcum* subsp. monococcum, 49 polymorphisms were identified, including 9 deletions, 10 insertions, 5 transitions and 9 transversions. It is interesting to note that there are common mutations in the chloroplast genomes of *T. sinskajae* and *T. monococcum* k-20970 that are absent in *T. monococcum* subsp. monococcum (Fig. 4). For example, ATAGC deletion at position 1503–1507 (Fig. 4a), TT insertion at position 8255–8256 (Fig. 4b), TTTT deletion at position 17,764–17,768 (Fig. 4c), AA insertion at position 105,192–105199 (Fig. 4d), TTC → GAA inversion at position 107,906–107908 (Fig. 4e) were found in both *T. sinskajae* and *T. monococcum* k-20970. It should be noted that the TTC sequence at position 107,906–107908 was present only in *T. monococcum* subsp. monococcum (Fig. 4e).

The comparison of the chloroplast genomes of *T. sinskajae* and *T. monococcum* k-20970 was of greatest interest, since their almost complete coincidence would indicate that one originated from another as a result of a spontaneous mutation. It was shown that the chloroplast genome of *T. sinskajae* differs from *T. monococcum* k-20970 in 10 nucleotides, including 1 deletion (Fig. 5a), 4 insertions (Fig. 5b, c, f, g), 1 inversion of AGAA to TTCT (Fig. 5d). An inversion occured in the *rbcL–psaI* intergenic region in the chloroplast genome of *T. monococcum* k-20970 (Fig. 5e), and in *T. monococcum* subsp. monococcum and *T. sinskajae*, this mutation was not detected.

The transversion A → T occured in the *trnL-UAA –trnF-GAA* intergenic region at position 47,765, A insertion in the *petN-trnC-GCA* intergenic region, T insertion in the *ycf4-cemA* intergenic region at position 59,120, A insertion in the *ndhE–ndhG* intergenic region at position 110,826, and T deletion in the *psbI-trnS* intergenic region at position 7469. All these mutations most likely do not affect the functions of the proteins, since they are located in intergenic regions. Insertion of an A at position 33,805 in

Table 2 Gene composition in chloroplast genome of *T. monococcum* k-20970 and *T. boeoticum* k-58674

Category of genes	Group of genes	Name of genes
Genes for photosynthesis	Subunits of ATP synthase	*atpA, atpB, atpE, atpF, atpH, atpI*
Genes for photosynthesis	Subunits of photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ, ycf3*
Genes for photosynthesis	Subunits of NADH-dehydrogenase	*ndhA, ndhB, ndhB, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
Genes for photosynthesis	Subunits of cytochrome b/f complex	*petA, petB, petD, petG, petL, petN*
Genes for photosynthesis	Subunits of photosystem I	*psaA, psaB, psaC, psaI, psaJ*
Genes for photosynthesis	Subunit of RuBisCO	*rbcL*
Self replication	Large subunit of ribosome	*rpl14, rpl16, rpl2, rpl2, rpl20, rpl22, rpl23, rpl23, rpl32, rpl33, rpl36*
Self replication	DNA dependent RNA polymerase	*rpoA, rpoB, rpoC1, rpoC2*
Self replication	Small subunit of ribosome	*rps11, rps12, rps12, rps14, rps15, rps15, rps16, rps18, rps19, rps19, rps2, rps3, rps3, rps4, rps7, rps7, rps8*
Other genes	c-type cytochrom synthesis gene	*ccsA*
Other genes	Envelop membrane protein	*cemA*
Other genes	Protease	*clpP*
Other genes	Translational initiation factor	*infA*
Other genes	Maturase	*matK*
Unkown	Conserved open reading frames	*ycf15, ycf15, ycf4**

[*] The duplicated *ycf15* gene was identified only in *T. monococcum* k-20970 and *T. boeoticum* k-58674, and *ycf4* was found in all einkorn wheats. The identification of additional ORFs with the same genome size is explained by the possibility of the coding genes being located in both nucleotide chains of chloroplast DNA

the *atpF* gene, responsible for ATP synthesis causes an open reading frame shift, which should affect the function of the protein.

Comparison of SSC, LSC, IRB, and IRB borders in the chloroplast genomes of einkorn wheats

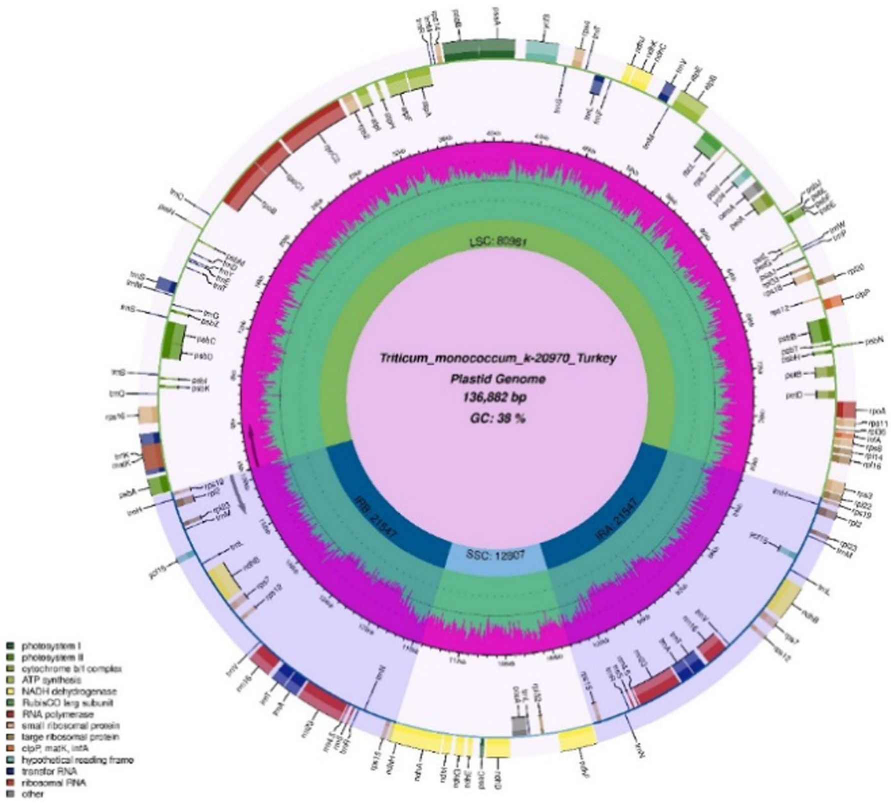
An analysis of the LSC–IRa, Ira–SSC, and SSC–IRb borders was carried out to identify differences between the chloroplast genomes of einkorn wheats. The border of LSC/IRb in einkorn wheats is located in the *psbA–rps19* intergenic region. In *T. sinskajae*, the size of the intergenic spacer is 27 bp before the border of this region and 51 bp after the border. *T. urartu* is the most diverged from other einkorn wheats by the border of LSC/IRb. The size of the *psbA–rps19* intergenic region in this species is 29 bp before the border and 39 bp after the border. *T. monococcum* and *T. sinskajae* were similar in this region of the chloroplast genome (Fig. 6a).

There were only minor differences between all analyzed einkorn wheat species at the border of Irb/SSC, which is located in the *rps15–ndhH* intergenic region of the chloroplast genome (Fig. 6b). As for the border of SSC/IRa, which is located in the *ndhF–rps15* intergenic region, *T. sinskajae, T. monococcum* subsp. monococcum*, T. monococcum* k-20970*, T. boeoticum* k-58674*,* and *T. urartu* were not polymorphic.

Phylogenetic analysis of einkorn wheats and some other species of the Triticum-Aegilops alliance based on comparison of chloroplast genomes

Sequencing and comparative analysis of nucleotide sequences showed the common origin of the chloroplast genomes of *T. sinskajae*, *T. boeoticum*, *T. monococcum* and *T. urartu*. However, some differences were identified between *T. sinskajae* and both samples of *T. monococcum*, which may indicate that these are two different species. According to the previously described analysis of the chloroplast genome and a resulting phylogenetic tree, *T. sinskajae* turned out to be closer to *T. monococcum* than to *T. boeoticum* and *T. urartu* (Fig. 7a). It was specifically close to the accession k-20970, from which it may have originated. In general, *T. sinskajae* and two samples

Fig. 2 Visual circular representation of the sequenced chloroplast genome of *T. monococcum* k-20970. Genes are shown in different colors; the purple area in the middle of the circle represents the GC content. IRA—inverted repeat region (**A**), IRB—inverted repeat region (**B**). Genes located outside the outer circle are transcribed clockwise, and genes located inside are transcribed counterclockwise



of *T. monococcum* turned out to be closely related and formed a separate cluster on the tree. *T. boeoticum* was more distant from *T. sinskajae*, and *T. urartu* was the most diverged (Fig. 7a).

Then, we compared the chloroplast genomes of einkorn and some polyploid wheats and representatives of *Aegilops*, using MAFFT v7.427 for alignment and Archaeopteryx JavaPlugin for construction of a phylogenetic tree (Fig. 7b). Chloroplast genome sequences of *T. sinskajae*, *T. monococcum* k-20970 and *T. boeoticum* k-58674 obtained in this study were aligned with chloroplast genome sequences of other wheat species from the GenBank (Table 1).

It is to note that the *Aegilops* species from the section Sitopsis belong to independent distant clades, one including a single species of *Ae. speltoides*, and another—*Ae. sharonensis*, *Ae. longissima*, *Ae. searsii*, *Ae. bicornis*. Actually, separation of two clades confirm the suggestion that this section consists of two subsections—Truncata and Emarginata, respectively (Miki et al. 2019). A separate branch is formed by *Ae. tauschii*, which is expected since this species comes from another section of Vertebrata. The chloroplast genomes of *Ae. speltoides* (both subspecies) and tetra- and hexaploid wheats from the timopheevii group represent two branches of the same clade. At the same time, *T. aestivum* L. and *T. turgidum* subsp. durum (Desf.) Husn. are distanced from them.

Discussion

T. sinskajae is a poorly studied einkorn wheat, probably originated as a result of a spontaneous mutation of *T. monococcum* (Filatenko and Kurkiev 2000). We earlier obtained new evidence that *T. sinskajae* can be considered a new species using classical molecular genetic analysis (Kuluev et al. 2018), and sequencing of chloroplast genomes (Kuluev et al. 2020). However, Sinskaya wheat is genetically compatible with *T. monococcum* and *T. boeoticum* (Kurkiev and Filatenko 2000). The combination *T. boeoticum* var. boeoticum x *T. sinskajae* has the highest crossability rate: the hybrid grain setting is 89%, the field germination of hybrid grains is 78.8%. In the hybrid combination *T. sinskajae* x *T.*

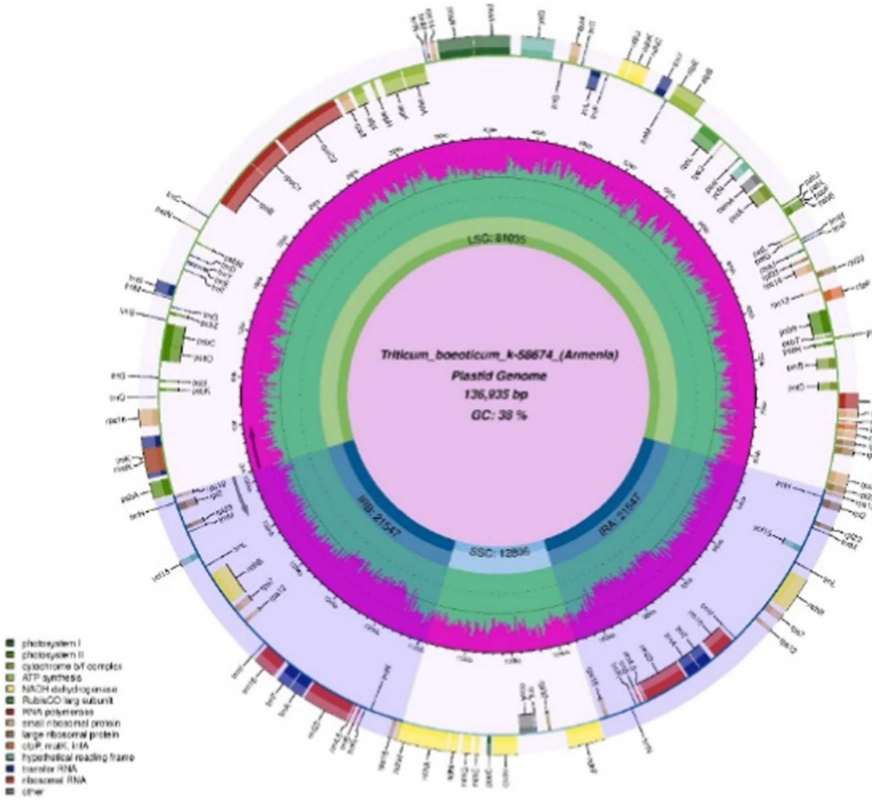


Fig. 3 Visual circular representation of the sequenced chloroplast genome of *T. boeoticum* k-58674. Genes are shown in different colors; the purple area in the middle of the circle represents the GC content. IRA—inverted repeat region (**A**), IRB—inverted repeat region (**B**). Genes located outside the outer circle are transcribed clockwise, and genes located inside are transcribed counterclockwise

Table 3 Main characteristics of the chloroplast genomes of *T. sinskajae* k-48993, *T. monococcum* k-20970, *T. boeoticum* k-58674, *T. monococcum* subsp. monococcum (LC005977) and *T. urartu* (KJ614411)

Genome characteristics	*T. sinskajae* k-48993	*T. monococcum* k-20970	*T. boeoticum* k-58674	*T. monococcum* subsp. *monococcum* (LC005977.1)	*T. urartu* (KJ614411.1)
Genome size (bp)	136,885	136,882	136,935	136,886	136,865
LSC length (bp)	80,983	80,981	81,035	80,986	80,948
SSC length (bp)	12,808	12,807	12,806	12,806	12,823
IR length (bp)	21,547	21,547	21,547	21,547	21,547
Genome length with one copy of the IR region (bp)	115,338	115,335	115,388	115,339	115,318
Number of different genes	130	132	132	130	130
Number of protein-coding genes	83	85	85	83	83
Number of tRNA genes	39	39	39	39	39
GC-content in LSC, %	36.24	36.24	36.24	36.24	36.25
GC-content in SSC, %	32.14	32.15	32.13	32.13	32.15
GC-content in IR, %	43.91	43.91	43.91	43.91	43.91
Total GC-content in the genome, %	38.27	38.27	38.27	38.27	38.28

monococcum var. hornemanii, the setting of hybrid grains was only 2.52%, indicating reproductive isolation, which could contribute to rapid speciation (Chelak and Chebotar 1983). However, it cannot be excluded that *T. sinskajae* originates from spontaneous mutagenesis in *T. monococcum* accession k-20970, where it was first discovered (Filatenko and Kurkiev 1975).

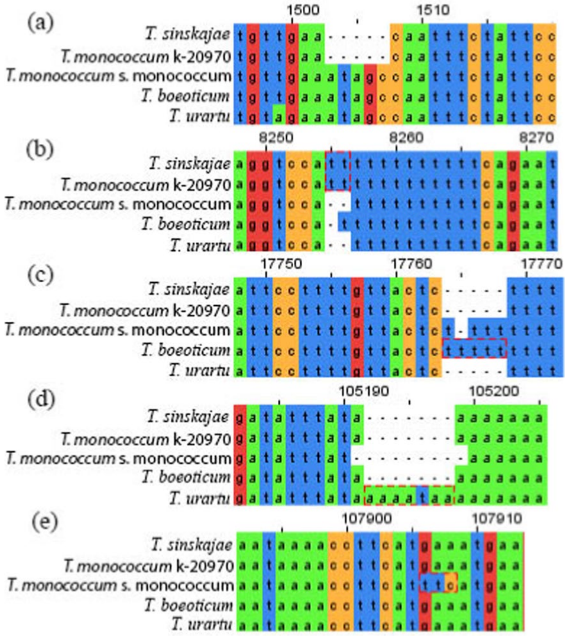


Fig. 4 Identical mutations detected in the chloroplast genomes of *T. sinskajae* and *T. monococcum* k-20970: **a** ATAGC deletion at position 1503–1507, **b** TT insertion at position 8255–8256, **c** TTTT deletion at position 17,764–17,768, **d** AA insertion in position 105,192–105199, **e** inversion GAA → TTC in position 107,906–107908. Nucleotide numbering is given according to the consensus sequence generated by the MAFFT v7.427 program

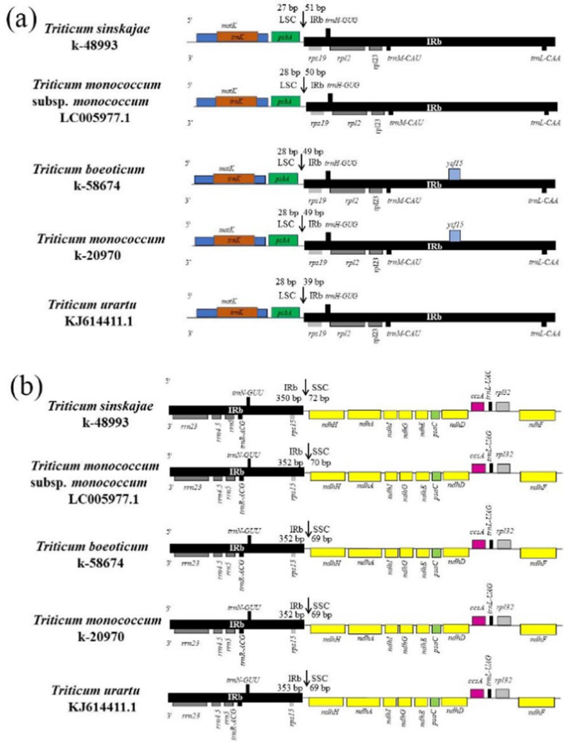


Fig. 6 Comparison of border positions of the LSC–IRb (**a**) and IRb–SSC (**b**) regions in the chloroplast genomes of *T. sinskajae*, *T. monococcum* subsp. monococcum, *T. monococcum* k-20970, *T. boeoticum* k-58674 and *T. urartu*

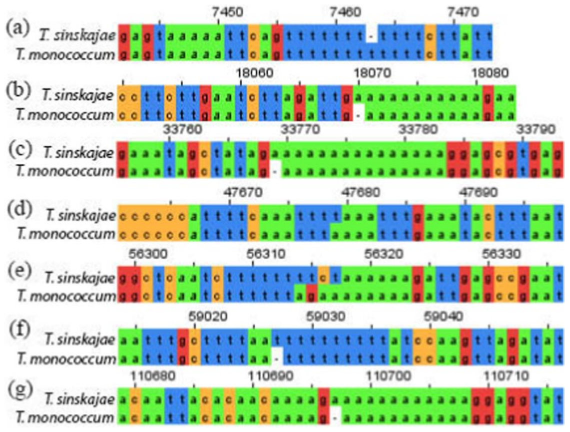


Fig. 5 Localization of mutations in the chloroplast genomes of *T. sinskajae* and *T. monococcum* k-20970. **a** deletion T at position 7463, **b** insertion A at position 18,070, **c** insertion A at position 33,769, **d** substitution of A with T at position 47,679, **e** inversion of AGAA to TTCT at position 56,314–563,117, **f** insertion T in position 59,027, **g** insertion A in position 110,697. Nucleotide numbering is given according to the consensus sequence generated by the MAFFT v7.427 program

Therefore, search for genetic differences between *T. sinskajae* and *T. monococcum* k-20970 is required to determine the status of Sinskaya wheat. Previously, such studies were carried out only using DNA markers. No differences between these wheats were detected in the chloroplast gene *matK* (Golovnina et al. 2007), but an insertion was found in the 11th intron of the nuclear *Acc-1* gene in *T. sinskajae* (Goncharov et al. 2007). One of the effective methods for phylogenetic analysis of *Triticeae* is the comparison of nucleotide sequences of complete chloroplast genomes (Gornicki et al. 2014). However, the chloroplast genome of *T. monococcum* k-20970 has not previously been sequenced. The deposited chloroplast genome of *T. boeoticum* (MG958551) has many long insertions and deletions that are absent in other einkorn wheats. Therefore, until now it has not been possible to perform a proper phylogenetic analysis of einkorn wheats. Within the framework of this study, the complete chloroplast genomes of *T. sinskajae*

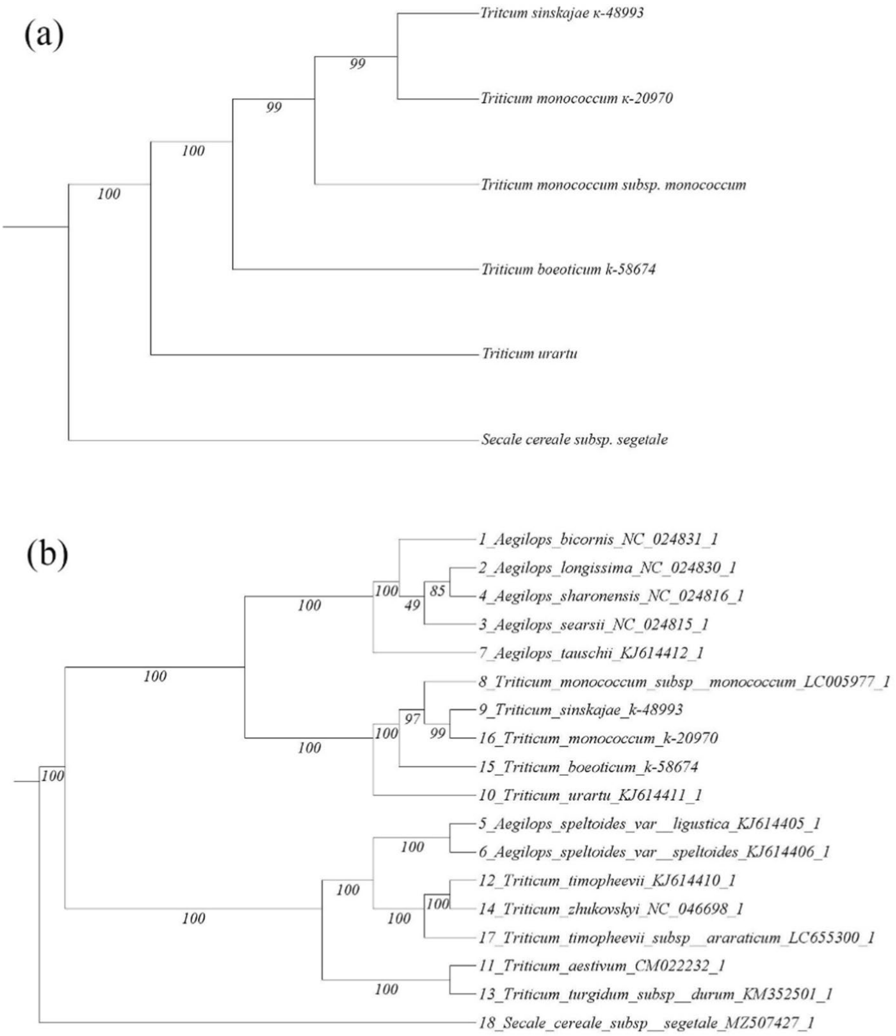


Fig. 7 **a** phylogenetic tree based on the alignment of chloroplast genomes of einkorn wheats *T. sinskajae*, *T. monococcum* subsp. monococcum (LC005977), *T. monococcum* (k-20970), *T. boeoticum* (k-58674) and *T. urartu* (KJ614411); **b** phylogenetic tree of some species of the Triticum-Aegilops alliance, forming the turgidum-aestivum and timopheevii lineages, built on the basis of comparison of the complete chloroplast genomes. In both trees, *Secale cereale* subsp. segetale (MZ507427.1) is represented as an outgroup species

k-48993, *T. monococcum* k-20970 and *T. boeoticum* k-58674 were sequenced and annotated for the first time. Comparison with data from GenBank made it possible to perform a phylogenetic analysis of all einkorn wheats. In general, all three chloroplast genomes were very similar in structure to previously published chloroplast genomes of einkorn wheats (KJ614411 and LC005977).

A more detailed analysis of the nucleotide sequences revealed differences between *T. sinskajae* and other einkorn wheats. *T. sinskajae* chloroplast genome had 48 polymorphisms compared to *T. monococcum* subsp. monococcum, and 245—compared to *T. urartu*. It is to note that only 10 polymorphisms were found in *T. sinskajae* compared to *T. monococcum* k-20970, including an inversion of 4 nucleotides,1 transversion, 1 deletion and 4 insertions. According to the results of a study of the chloroplast genome, *T. sinskajae* and *T. monococcum* have a common maternal origin and are closely related. However, several deletions, insertions and substitutions in the chloroplast genome of *T. sinskajae* suggest that its unique phenotype is not a result of a single spontaneous mutation of *T. monococcum*. These two wheats most probably evolved in reproductive isolation for a long time, but within the same area. There is no doubt that *T. sinskajae* separated from the whole branch of einkorn wheats or from *T. monococcum* later than other einkorns, but nevertheless managed to accumulate quite a significant number of mutations both in the nuclear (Goncharov et al. 2007; Watanabe 2017; Kuluev et al. 2018) and in the chloroplast genomes. Representatives of the genus *Aegilops* are considered different species despite the similarity of chloroplast genomes and morphology. Comparison of plastomes in Emarginata subsection (from the Sitopsis section) revealed that there were very few differences between species. *Ae. bicornis* differs from *Ae. searsii*, *Ae. longissima* and *Ae. sharonensis* by 27 nucleotide substitutions. *Ae. searsii* differs from *Ae. longissima* and *Ae. bicornis* by 27 substitutions, and by 28 substitutions—from *Ae. sharonensis*. *Ae. longissima* chloroplast genome has 12 substitutions compared to *Ae. sharonensis*, and 17 substitutions compared to *Ae. searsii* and *Ae. bicornis*. *Ae. sharonensis* differs from *Ae. longissima* by 23 substitutions, from *Ae. bicornis*—by 26 substitutions, and from *Ae. searsii* by 27 substitutions. It should be noted that other types of mutations were not detected between chloroplast genomes of representatives of Emarginata subsection. Despite a few number of nucleotide substitutions and 99.98–99.99% homology of the chloroplast genomes sequences, they are considered distinct species (Kuluev et al. 2020; Miki et al. 2019).

Therefore, 10 mutations of different types we identified in *T. sinskajae*, compared to *T. monococcum* accession k-20970, indicate distinct species status of Sinskaya wheat. Nevertheless, 49 polymorphisms were found between *T. monococcum* subsp. monococcum and *T. monococcum* accession k-20970. According to this data, *T. monococcum* accession k-20970 and *T. sinskajae* are equally distant from *T. monococcum* subsp. monococcum. Therefore, these three einkorn wheats could also be considered as subspecies of *T. monococcum*. It should be noted that *T. monococcum* has a large number of lineages, some of which should be elevated to subspecies status. In this regard, status revision of numerous *T. monococcum* lineages should be continued using genetic methods such as sequencing and analysis of chloroplast genomes. In general, the results of phylogenetic analysis support clear identification of distinct species—*T. boeoticum, T. monococcum* and the most diverged *T. urartu* (Fig. 7). Obviously, *T. urartu* was the earliest to separate from the general evolutionary lineage of einkorn wheat.

We also carried out phylogenetic analysis of some species of the Triticum-Aegilops alliance from the turgidum-aestivum and timopheevii lineages. Our results confirm that *Aegilops* species from the Sitopsis section belong to independent clades, one represented by the only species *Ae. speltoides*, and another—by *Ae. sharonensis*, *Ae. longissima*, *Ae. searsii*, *Ae. bicornis*. These two clades of *Aegilops* belong to two different subsections—Truncata and Emarginata, respectively (Miki et al. 2019). *Ae. tauschii* is slightly diverged (Fig. 7) from the subsection Emarginata, because this *Aegilops* is considered to belong to another section—Vertebtata (Wang et al. 1997). Chloroplast genomes of *Ae. speltoides* (both subspecies) and tetra- and hexaploid wheat from the timopheevii lineages belong to different branches of the same clade. Minor differences between *T. timopheevii* and *T. zhukovskyi* support the suggestion that the first species was the maternal ancestor of the second (Menabde and Yeritsyan 1960).

The wheats of the turgidum-aestivum lineages are further distanced, which may indicate that the donor of subgenome B for these wheats was not *Ae. speltoides*, but a closely related species or subspecies of *Ae. speltoides*, which is probably extinct.

For the past 100 years, the search for donors of subgenomes B, A and D of bread wheat has been of great interest to researchers, which allowed to clarify phylogenetic relationships among species of the genera *Triticum* and *Aegilops*. Among all these species, *T. sinskajae* occupies a special position. For many years *T. sinskajae* was considered a naked spontaneous mutant of *T. monococcum*, and not a species. However, our results demonstrate that *T. sinskajae* is genetically different from *T. monococcum*. It is the only free-threshing species among einkorn wheats. Its phenotypic properties, such as a semi-compact spike, soft ear glumes which allow easy threshing, and the highest protein content among einkorns, indicate that this wheat is somewhat unique. *T. sinskajae* differs from *T. monococcum* in its more compact ear shape and shorter stem. All these traits could not result from one or even several spontaneous mutations, but most likely are an evolutionary acquisition of this species of wheat. Rapid speciation could occur in *T. sinskajae* as a result of reproductive isolation from *T. monococcum* (Chelak and Chebotar 1983), however, additional research is required to confirm this hypothesis. In this regard, it is of great interest to sequence and compare the nuclear genomes of *T. sinskajae* and *T. monococcum* k-20970 to obtain more evidence of the species identity of Sinskaya wheat.

T. sinskajae differs from *T. monococcum* in its more compact ear shape and shorter stem height. These traits cannot be the result of one or even several spontaneous mutations, but most likely are an evolutionary acquisition of this species of wheat. The reason for rapid speciation in the case of *T. sinskajae* may be the resulting reproductive isolation with *T. monococcum* (Chelak and Chebotar 1983), however, additional research is required to confirm this hypothesis. In this regard, it is of great interest to sequence and compare the nuclear genomes of *T. sinskajae* and *T. monococcum* k-20970, which will make it possible to put a final point on the question of the species identity of the first of them. The chloroplast genome sequences were deposited in GenBank, accession numbers: *T. sinskajae* (k-48993) OR803873, *T. monococcum* (k-20970) OR936050, *T. boeoticum* (k-58674) PP067985.

Acknowledgements The study of A.R. Kuluev was supported by state assignment no. 122030200143-8. The authors express their deep gratitude for the seeds used in the study, which were kindly provided from the world collection of seeds of the Federal Research Center "N.I. Vavilov All-Russian Institute of Plant Genetic Resources" (VIR, St. Petersburg).

Author contributions Bulat Kuluev and Alexey Chemeris designed research. Azat Kuluev isolated chloroplast DNA and performed sequencing. Bulat Kuluev conducted phylogenetic analysis. Azat Kuluev, Bulat Kuluev and Alexey Chemeris interpreted the results, wrote the paper and participated in the editing of the article. Elena Mikhaylova carried out a professional technical English edit of the manuscript.

Funding The study of B.R. Kuluev and E.V. Mikhaylova was funded by the Ministry of Science and Higher Education of the Russian Federation (agreement no. 075-15-2021-1066 dated September 28, 2021).

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bernhardt N, Brassac J, Kilian B, Blattner F, Bernhardt N, Brassac J, Blattner F (2017) Dated tribe-wide whole chloroplast genome phylogeny indicates recurrent hybridizations within Triticeae. BMC Evol Biol 17:141. https://doi.org/10.1186/s12862-017-0989-9

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bosacchi M, Gurdon C, Maliga P (2015) Plastid genotyping reveals uniformity of cytoplasmic male sterile-t maize cytoplasms. Plant Physiol 169:2129–2137. https://doi.org/10.1104/pp.15.01147

Chelak VR, Chebotar AA (1983) Cytogenetics of *Triticum sinskajae* A. Filat. et Kurk. and its interspecific hybrids. Cytol Genet 17(1):21–26

Dorofeev VF, Filatenko AA, Migushova EF (1979) Cultural flora of the USSR. Wheat. Kolos, Leningrad

Farrar K, Donnison I (2007) Construction and screening of BAC libraries made from Brachypodium genomic DNA. Nat Protoc 2(207):1661–1674. https://doi.org/10.1038/nprot.2007.204

Filatenko AA, Kurkiev UK (1975) Sinskaya wheat (New species—*Triticum sinskajae* A. Filat. et Kurk.). Proceed Appl Bot Genet Breed 54(1):239–241

Gogniashvili M, Maisaia I, Kotorashvili A, Kotaria N, Beridze T (2018) Complete chloroplast DNA sequences of Georgian indigenous polyploid wheats (*Triticum* spp.) and B plasmon evolution. Genet Resour Crop Evol 65:1995–2002. https://doi.org/10.1007/s10722-018-0671-0

Golovnina KA, Glushkov SA, Blinov AG, Mayorov VI, Adkison LR, Goncharov NP (2007) Molecular phylogeny of genus Triticum L. Plant System Evol 264:195–216. https://doi.org/10.1007/s00606-006-0478-x

Golovnina KA, Kondratenko EY, Blinov AG, Goncharov NP (2009) Phylogeny of A-genomes of wild and cultivated wheat species. Russian Jour Genet 45(11):1540–1547. https://doi.org/10.1134/S1022795409110106

Goncharov NP, Kondratenko EJ, Bannikova SV, Konovalov AA, Golovnina K (2007) Comparative genetic analysis of diploid naked wheat *Triticum sinskajae* and the progenitor *T. monococcum* accession. Rus J Genet 43:1248–1256. https://doi.org/10.1134/S1022795407110075

Gornicki P, Zhu H, Wang J, Challa GS, Zhang Z, Gill BS, Li W (2014) The chloroplast view of the evolution of polyploid wheat. New Phytol 204(3):704–714. https://doi.org/10.1111/nph.12931

Graham DE (1978) The isolation of high molecular weight DNA from whole organisms of large tissue masses. Anal Biochem 78:673–678

Han MV, Zmasek CM (2009) phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinform 10:356. https://doi.org/10.1186/1471-2105-10-356

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30(4):772–780. https://doi.org/10.1093/molbev/mst010

Kuluev AR, Matnijazov RT, Kuluev BR, Chemeris AV (2018) A molecular genetic research of the *Triticum sinskajae* A. Filat. et Kurk. by RAPD analysis and by comparing the nucleotide sequences of the variable intergenic region of the petN-trnC-GCA chloroplast genome and intron of the histone H3.2 gene. Ecol Genet 16(1):53–59

Kuluev AR, Matniyazov RT, Chemeris DA, Yunusbaev UB, Kuluev BR, Chemeris AV (2020) Phylogenetic relationships in the wheat-aegilops alliance through the prism of the chloroplast genome. Biomics 12(4):532–544

Kurkiev UK, Filatenko AA (2000) New forms of Sinskaya wheat (*Triticum sinskajae* A.Filat et Kurk.) with light grain threshing and stunting genes. Rep Russian Agricult Academ Science 4:10–12

Menabde VL, Yeritsyan AA (1960) To the study of Georgian wheat Zanduri. Message Academ Science Geor SSR 25:731–736

Miki Y, Yoshida K, Mizuno N, Nasuda S, Sato K, Takumi S (2019) Origin of wheat B-genome chromosomes inferred from RNA sequencing analysis of leaf transcripts from section Sitopsis species of *Aegilops*. DNA Res 26(2):171–182. https://doi.org/10.1093/dnares/dsy047

Nie X, Zhao X, Wang S, Zhang T, Li C, Liu H, Tong W, Guo Y (2018) Complete chloroplast genome sequence of Broomcorn millet (*Panicum miliaceum* L.) and comparative analysis with other Panicoideae species. Agronomy 8(159):1–18. https://doi.org/10.3390/agronomy8090159

Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C (2019) CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res 47:W65–W73. https://doi.org/10.1093/nar/gkz345

Simons KJ, Fellers JP, Trick HN, Zhang Z, Tai YS, Gill BS, Faris JD (2006) Molecular characterization of the major wheat domestication gene Q. Genetics 172(1):547–555. https://doi.org/10.1534/genetics.105.044727

Su Q, Liu L, Zhao M, Zhang C, Zhang D, Li Y, Li S (2020) The complete chloroplast genomes of seventeen *Aegilops tauschii*: genome comparative analysis and phylogenetic inference. PeerJ 8(e8678):1–19. https://doi.org/10.7717/peerj.8678

Vavilova VYu, Konopatskaya ID, Blinov AG, Goncharov NP (2020) Evolution of the Btr1-A gene in diploid wheat species of the genus Triticum L. Rus J Genet 56(5):609–614. https://doi.org/10.1134/S1022795420050142

Wang GZ, Miyashita NT, Tsunewaki K (1997) Plasmon analysis of Triticum (wheat) and *Aegilops*: PCR–single-strand conformational polymorphism (PCR-SSCP) analyses of organellar DNAs. Proc Natl Acad Sci USA 94:14570–14577. https://doi.org/10.1073/pnas.94.26.14570

Watanabe N (2017) Breeding opportunities for early, free-threshing and semidwarf *Triticum monococcum* L. Euphytica 213:201. https://doi.org/10.1007/s10681-017-1987-0

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25(9):1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Wu P, Xu C, Chen H, Yang J, Zhang X, Zhou S (2021) NOVOWrap: an automated solution for plastid genome assembly and structure standardization. Mol Ecol Resour 21:2177–2186. https://doi.org/10.1111/1755-0998.13410

Zhang Y-J, Ma P-F, Li D-Z (2011) High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody Bamboos (Poaceae: Bambusoideae). PLoS ONE 6(5):e20596. https://doi.org/10.1371/journal.pone.0020596

Zheng S, Poczai P, Hyvönen J, Tang J, Amiryousefi A (2020) Chloroplot: an online program for the versatile plotting of organelle genomes. Front Genet 11:576124. https://doi.org/10.3389/fgene.2020.576124