RESEARCH

Complete mitochondrial genome of *Agropyron cristatum* reveals gene transfer and RNA editing events

Taiyou Ou[1,2], Zinian Wu[1,2*], Chunyu Tian[1,2], Yanting Yang[1,2] and Zhiyong Li[1,2]

Abstract

Background As an important forage in arid and semi-arid regions, *Agropyron cristatum* provides livestock with exceptionally high nutritional value. Additionally, *A. cristatum* exhibits outstanding genetic characteristics to endure drought and disease. Therefore, rich genetic diversity serves as a cornerstone for the improvement of major food crops. The purposes of this study were to systematically describe mitogenome of *A.cristatum* and preliminarily analyze its internal variations.

Result The *A. cristatum* mitogenome was a single-ring molecular structure of 381,065 bp that comprised 52 genes, including 35 protein-coding, 3 rRNA and 14 tRNA genes. Among these, two pseudoprotein-coding genes and multiple copies of tRNA genes were observed. A total of 320 repetitive sequences was found to cover more than 10% of the mitogenome (105 simple sequences, 185 dispersed and 30 tandem repeats), which led to a large number of fragment rearrangements in the mitogenome of *A. cristatum*. Leucine was the most frequent amino acid ($n = 1087$, 10.8%) in the protein-coding genes of *A. cristatum* mitogenome, and the highest usage codon was ATG (initiation codon). The number of A/T changes at the third base of the codon was much higher than that of G/C. Among 23 PCGs, the range of Pi values is from 0.0021 to 0.0539, with an average of 0.013. Additionally, 81 RNA editing sites were predicted, which were considerably fewer than those reported in other plant mitogenomes. Most of the RNA editing site base positions were concentrated at the first and second codon bases, which were C to T transitions. Moreover, we identified 95 sequence fragments (total length of 34, 343 bp) that were transferred from the chloroplast to mitochondria genes, introns, and intergenic regions. The stability of the tRNA genes was maintained during this process. Selection pressure analysis of 23 protein-coding genes shared by 15 Poaceae plants, showed that most genes were subjected to purifying selection during evolution, whereas *rps4*, *cob*, *mttB*, and *ccmB* underwent positive selection in different plants. Finally, a phylogenetic tree was constructed based on 22 plant mitogenomes, which showed that *Agropyron* plants have a high degree of independent heritability in *Triticeae*.

Conclusion The findings of this study provide new data for a better understanding of *A. cristatum* genes, and demonstrate that mitogenomes are suitable for the study of plant classifications, such as those of *Agropyron*. Moreover, it provides a reference for further exploration of the phylogenetic relationships within *Agropyron* species,

*Correspondence:
Zinian Wu
wuzinian@caas.cn

Full list of author information is available at the end of the article

and establishes a theoretical basis for the subsequent development and utilization of *A. cristatum* plant germplasm resources.

Keywords *Agropyron cristatum*, Gene transfer, Mitogenome, RNA editing, *Triticeae*

Background

Triticeae is a botanical tribe that comprises a large number of grass species that can be used as a mainstay of food and animal husbandry. In particular, crested wheatgrass complex (*Agropyron* spp.), which can be found in mountains, grassland, deserts, and other areas of Northern Europe and Asia [1–3]. It has a wide ecological range, so that performs excellently in low temperature, water deficiency and other adverse adaptations. Furthermore, it has outstanding advantages in controlling weed invasion and creating a conducive habitat for livestock owing its perennial and strong root growth [3, 4]. *Agropyron* are polyploid plants that were often erroneously grouped with plants of other genera within *Triticaae* during its early classification [1]. Analysis of its chromosomes revealed the presence of P-type genome that is genome symbol (a diploid *Agropyron* species is shown as P), which allowed to formally establish its classification [5]. At present, 10~15 species of *Agropyron* species have been identified (exclude infraspecific names), some of which are controversial. There are 5 species (1 endemic species) in China (*A.cristatum, A.desertorun, A.michnoi, A.sibiricum, A.mongolicum*) [1, 2, 5].

Chloroplasts and mitochondria in plants are extremely important semi-autonomous organelles, and their genomes are of great significance for research. Since the publication of the first mitochondrial genome in 1981 [6], subsequent research has clarified the mitochondrial structure and its changing features in different species. Plant mitogenomes are larger length and have more complex variations than those in animals [7–9]. Most plant mitogenomes range in size from 200 to 800kb [10], with the two most extreme examples found to data being those of spermatophytes of *Viscum scurruloideum* (66 kb) [11] and *Larix sibirica* (11.7 Mb) [12]. Initially, the structure of plant mitochondria was generally recognized as a mainly circular; however, further studies revealed that mitochondria also have non-circular structures (i.e., circularly permuted linear molecules) [8], which are a consequence of the frequent recombination of a large number of repetitive sequences, insertion of exogenous genes and loss of internal genes in plant mitochondria [13, 14]. Under such high frequency of mutations, the mitochondria lacks a DNA repair system, resulting in more complex structural changes in plant mitochondria [15]. Paradoxically, mitochondria undertake the key task of supplying energy and are involved in various physiological activities, including male sterility in plant cells [16]. As one of the primary carriers of complete genetic information within a cell, the mitogenome is also a very important pathway for studying the genetic evolution of species themselves.

The initial traditional classification system relying on morphological phenotypes was characterized by its generality and broad scope, owing to which *Agropyron* was once considered the largest genus in the *Triticeae* tribe (with approximately 100 plant species) [17]. Subsequently, cytogenetic analysis helped to accurately determine the chromosome types of the *Agropyron* genus and define a recognizable group of plants. In recent years, researchers have identified the chromosomes in diploid *Agropyron* using fluorescent in situ hybridization analysis, which has allowed to establish the phylogenetic relationships between *Agropyron* spp. and hexaploid wheat [18]. Furthermore, analysis of different molecular markers, such as simple sequence repeats (SSRs) and gliadin, as well as the identification of several single nucleotide polymorphisms (SNPs) within the 214,854 transcript sequences recognized via high-throughput sequencing have provided a detailed description of the genetic relationships between species of the *Agropyron* genus and the *Triticeae* tribe [19–21]. Notably, whole-genome phylogenetic analysis revealed that the *Agropyron* genus has an independent genetic position within the *Triticeae* tribe [22]. Accumulated molecular evidences suggest that *Agropyron* spp. have valuable genetic features that can be used to improve the ecological and economic output of other forage grasses and major food crops. Therefore, a better understanding of the evolutionary classification of the *Agropyron* genus may facilitate the development of new approaches to explore the genetic features of *Agropyron* spp. more comprehensively.

The present study aimed to characterize the mitogenome of *A. cristatum* and provide insights into the genetic variation events that have occurred throughout its evolution. The collected genetic data is expected to provide an effective theoretical basis and reference for the genetic analysis and subsequent utilization of *Agropyron* features for greater societal and economic impacts.

Materials and methods

DNA extraction, genome sequencing, and assembly

Fresh leaves of *A. cristatum* were collected from Hohhot, Inner Mongolia, China (40.57°N, 111.93°E) and deposited in the National Medium-Term Genebank Forage Germplasm (Hohhot, China). Genomic DNA was extracted from the leaves using a Plant DNA Isolation Kit (Tiangen Biotech, Beijing, China) and sequenced using Novaseq6000 (Illumina, San Diego, CA, USA)

Ou *et al. BMC Plant Biology*   (2024) 24:830

Page 3 of 17

and PromethION (Oxford Nanopore Technologies, Oxford, UK) sequencing systems. PromethIONs data were aligned to the reference gene sequence using the minimap2 (v2.1) algorithm [23] to obtain the mitogenome sequence and was corrected using canu (v2.2) [24]. Bowtie2 (v2.3.5.1) [25] was used to align the second-generation sequencing data with the corrected sequences. Subsequently, the aligned second-generation and corrected third-generation data were assembled using Unicycler (v0.4.8) [26] with default parameters. Finally, the assembly results are visualized and manually adjusted using Bandage software (v0.8.1) [27].

Genome annotation

The mitogenome of *A.cristatum* was annotated using GeSeq [28] with reference to previously released mitogenome data of *Triticeae* species (accession number: NC_036024.1, KJ078649.1 and NC_022714.1) and was then manually adjusting into a circular mitogenome model. The chloroplast genome was conducted using the second-generation raw data from the same material, assembled with the software GetOrganelle (v1.7.7.0) [29], and annotated with the PGA tools [30]. The genome map was visualized using the Organellar Genome Draw (OGDRAW) software (v1.3.1) [31].

Repeat sequence identification

A. cristatum mitogenome was analyzed using MISA [32] to identify SSR motifs with lengths of at least mono-10, di-5, tri-4, tetra-3, penta-3, and hexa-3. Tandem Repeats Finder [33] was used with standard settings for tandem repeat identification. REPuter [34] was used to identify dispersed repeats exceeding 29 base pairs and using a Hamming distance threshold of 3 and an E-value cutoff of 1×10^{-5} to classify them as forward, reverse, palindromic, or complementary repeats.

Codon usage bias analysis

The relative synonymous codon usage (RSCU) values of the protein-coding genes (PCGs) and their amino acid compositions were calculated using the CodonW software, and codon preferences were configured using Perl scripts. to select unique CDS, and plotted the results with R.

Selective pressure calculation and nucleotide diversity (pi) analysis

Non-synonymous (Ka) and synonymous (Ks) substitution rates were calculated using DnaSP (v6.12.0) [35], based on a total of 23 shared PCGs—*atp1, atp4, atp6, atp8, atp9, ccmB, ccmFc, ccmFn, cob, cox1, cox2, cox3, mttB, nad4, nad6, nad7, nad9, rps1, rps2, rps4, rps7, rps13* and *rps14*—within the mitogenomes of *A.cristatum* and 15 other Poaceae spp. that sequence alignment performed

by MAFFT (version 7.131) [36]. We subsequently calculated the Pi values for each gene with DnaSP (v6.12.0).

Prediction of RNA editing sites

Based on three RNA-sequencing datasets of *A. cristatum* deposited in the Sequence Read Archive database (https://www.ncbi.nlm.nih.gov/sra/; accession numbers: SRR22796184, SRR22796185 and SRR22796186), we identified the putative RNA editing sites in mitochondrial PCGs. RNA sequences of mitochondrial PCGs were aligned using BWA software (v0.7.15) [37]. Subsequently, SAMtools (v1.17) [38] and BCFtools (v1.17) [39] were employed for SNP identification. For RNA editing site detection and annotation, SNP data were analyzed with REDO (v1.0) [40] in default parameters. To filter out spurious RNA-editing sites, DNA-seq data were mapped to the *A. cristatum* mitogenome using BWA and BCFtools were applied to identify genomic SNPs, thereby excluding any overlapping sites from the RNA editing candidates.

Chloroplast-derived mitochondrial sequence identification

The chloroplast genome sequences of *A. cristatum* were derived from in-house assemblies. To identify homologous segments between the chloroplast and mitochondrial genomes, BLASTN was used with an *E*-value threshold of 1×10^{-5} to screen for the transferred DNA fragments. The transfer of genes from the chloroplasts to the mitochondria was visually depicted using TB tools (v2.091) [41].

Phylogenetic analysis

Phylogenetic analysis was performed based on 23 shared PCGs derived from the complete mitogenomes of 20 selected Poaceae spp., with *Glycine max* and *Medicago truncatula* as the outgroups. The mitogenome data for the reference accessions were obtained from the National Center for Biotechnology Information database (Table S1). Nucleotide sequences of the selected PCGs were concatenated and multiple sequence alignments were performed using MAFFT (v7.131) [36]. Phylogenetic tree construction involved implementing Maximum Likelihood (ML) with optimal models determined using ModelFinder (IQ-Tree v2.3.2) [42]. Specifically, the ML tree was inferred via RAxML [43] using the GTRGAMMA model supported by 1,000 bootstraps.

Results
Genomic features of the *agropyron cristatum* mitogenome

Based on our assembly results, the mitogenome structure of *A. cristatum* was the same as that of most plants. It had a single circular molecular structure with a length of 381,065 bp (Fig. 1) and a GC content of 44.28%. In the complete mitochondrial genome of *A. cristatum*, we have identified and annotated 52 genes. According to their
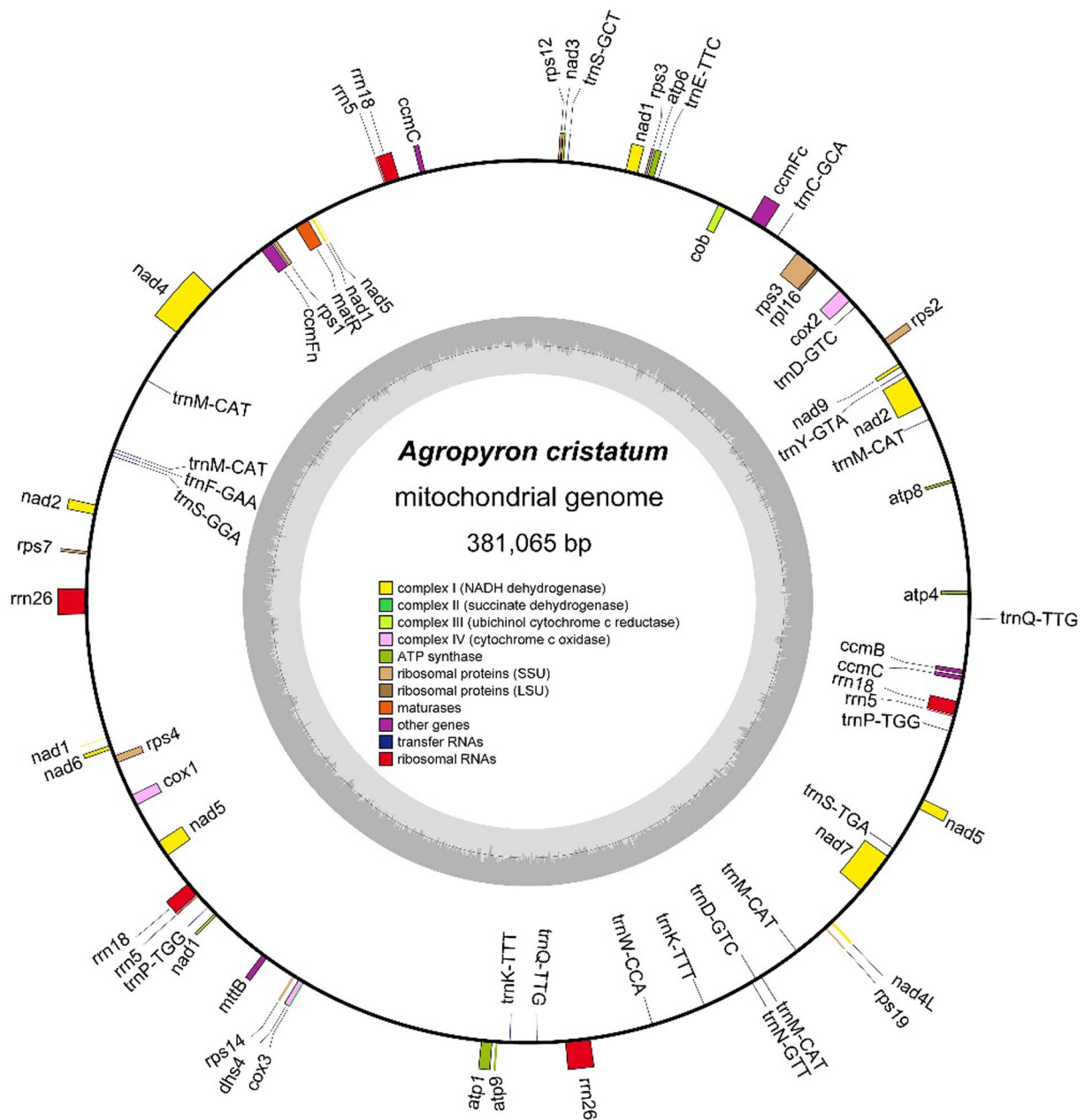
Fig. 1 Circular map of the complete mitogenome of *Agropyron cristatum*. Genes indicated outside and inside of the circle are forward and reverse transcribed, respectively. The dark-grey region in the inner circle depicts the GC content. Different functional gene groups are represented with different colors

different functions, they were further divided into 4 main groups (Table 1). The first group contained 14 core genes, and the second group contained 21 variable genes, 50% of which were related to Ribosomal proteins. These 35 sequences were PCGs in the mitogenome that belonged to ten different functional groups. 3 rRNA genes and 14 tRNA genes in another two groups. In the process of constructing genome structure and annotating genes (Table S2), we found that *ccmFc, cox2, rps1* and *rps3* all contained one intron; *nad4* contained three introns; and

nad1, nad2, nad5 and *nad7* have four introns in these 52 genes. The multi-copy genes were mainly concentrated in the two groups of rRNA and tRNA. Among them, *rrn26, trnD-GTC, trnK-TTT, trnP-TGG* and *trnQ-TTG* owned two copies; *rrn18* and *rrn5* owned three copies; while *trnM-CAT* owned five copies. In addition, only *ccmC* was a protein-coding gene with two copies. Two pseudogenes (*rps14* and *sdh4*) were also identified among the annotated mitochondrial genes.

Ou *et al. BMC Plant Biology*   (2024) 24:830

Page 5 of 17

Table 1 Gene composition in the *Agropyron cristatum* mitogenome

	Group of genes	Gene name
Core genes	ATP synthase	atp1, atp4, atp6, atp8, atp9
	Cytochrome c biogenesis	ccmB, ccmC(2), ccmFc*, ccmFn
	Ubiquinol cytochrome c reductase	cob
	Cytochrome c oxidase	cox1, cox2*, cox3
	Maturases	matR
Variable genes	Transport membrane protein	mttB
	NADH dehydrogenase	nad1****,nad2****,nad3,nad4***,nad 4L, nad5****,nad6,nad7****,nad9
	Ribosomal proteins (LSU)	rpl16
	Ribosomal proteins (SSU)	#rps14,rps1*,rps12,rps13,rps19,rps2,rps 3*,rps4,rps7
	Succinate dehydrogenase	#sdh4
rRNA genes	Ribosomal RNAs	rrn18(3), rrn26(2), rrn5(3)
tRNA genes	Transfer RNAs	trnC-GCA, trnD-GTC(2), trnE-TTC, trnF-GAA, trnK-TTT(2), trnM-CAT(5), trnN-GTT, trnP-TGG(2) ,trnQ-TTG(2), trnS-GCT, trnS-GGA, trnS-TGA, trnW-CCA, trnY-GTA

* Represent introns. # Indicate pseudogenes. Numbers within parentheses indicate the number of copies of multicopy genes

Anatomization of repeat sequence

While analyzing the repeat sequences in *A. cristatum* mitogenome, we mainly focused on three types: simple sequence, tandem, and dispersed repeats (Fig. 2). A total of 97 complete and 4 compound SSRs sequences were identified in the entire *A. cristatum* mitogenome (Table S3). After further splitting, 105 SSRs were obtained (Fig. 3). Among the 1–6 nucleotide repeat types, the number of tetranucleotide repeats accounted for the largest proportion (34.3%), followed by mononucleotide (26.7%), dinucleotide (16.2%), trinucleotide (13.3%), pentanucleotide (7.6%), and hexanucleotide (1.9%). From sizes of the repeat type, their coverage on the complete mitogenome is not extensive that the highest coverage is only 0.1%. A total of 32 different repeat motifs (Table S4) were identified among all SSRs, the average motif counts for different types of repeats, from high to low, is as follows: mononucleotide ($n=14$), dinucleotide ($n=5.7$), tetranucleotide ($n=3$), trinucleotide ($n=2.3$), hexanucleotide ($n=2$), pentanucleotide ($n=1$). From the all motifs, the most frequent repeat motif being of A/T ($n=24$, 22.9%), followed by the sequence with AATG/AATC ($n=16$, 15.2%) and AT/AT ($n=10$, 9.5%) as the repeat motifs. The number of other repeat motifs did not exceed 5 (4.8%). Although these sequences were composed of different repeat types and motifs, their distributions in the mitogenome were highly similar. Among the 105

SSRs, 88.6% ($n=93$) of the sequences were located in the intergenic spacer (IGS), 6.7% ($n=7$) in the introns, and 4.7% ($n=5$) in the coding regions of the five genes (*atp8*, *rps2*, *rps3*, *ccmFn*, and *nad1*).

Another type of repetitive sequence appeared in *A. cristatum* mitogenome was the dispersed repeats. Using the corresponding parameter settings, a total of 185 (11.5%) dispersed repetitive sequences with a mean length of approximately 29 bp (total length of 43,916 bp) were detected (Fig. 4 and Table S5). Among them, 92 (49.7%) were forward repeat sequences and 93 (50.3%) were palindromic repeat sequences. Although the two types of repeat sequences were evenly distributed within the mitogenome, they were differently distributed considering their lengths. Five sequences exceeded 2 kb, with the longest forward repeat sequence being 8,337 bp that encoded two *rrn26* genes, and the longest palindromic repeat sequence being 7,814 bp that encoded six genes (two *rrn18*, two *ccmC*, and two *rrn5*). The length distribution of scattered repeat sequences in the *A. cristatum* mitogenome could be roughly divided into three regions, with the main distribution area ranging from 29 to 69 bp, of which the number of sequences ranging from 40 to 49 bp was the largest (forward: 27, palindromic: 27; 29.2%), whereas the distribution area ranging from 70 to 99 bp was the smallest (forward: 15, palindromic: 13; 15.1%). In addition, sequences longer than 100 bp also had a certain number (forward: 18, palindromic: 12; 16.2%). The proportion of sequences that were completely or partially located in the IGS region among the scattered repeat sequences was 86.5%, which was similar to the distribution of microsatellite repeat sequences.

In addition, we searched for 30 tandem repeat sequences in the complete mitogenome of *A. cristatum*, with a length range of 6 to 70 bp and a matching degree of 78% or higher. Fifteen sequences (50%) had a matching degree of 100% (Table S6). The copy number variation of these 30 tandem repeat sequences ranged between 1.9 and 4.3, indicating that a considerable portion of the tandem repeat sequences were incomplete copies. Unsurprisingly, the distribution of the identified tandem repeat sequences within the mitogenome followed the same trend as that of the other two types of repetitive sequences: 28 sequences (93.3%) were distributed in the IGS region of each gene, and only two tandem repeat sequences were located in the coding regions of *rps2* and *rps4*.

Codon usage analysis of PCGs

In total, 10,101 codons were used to encode 35 PCGs in the *A. cristatum* mitogenome (Table 2). The top three most frequently amino acids were Leu (10.8%), Ser (9.0%), and Ile (7.9%), whereas Ter was the least frequent amino acid with only 33 occurrences (0.3%). Other amino
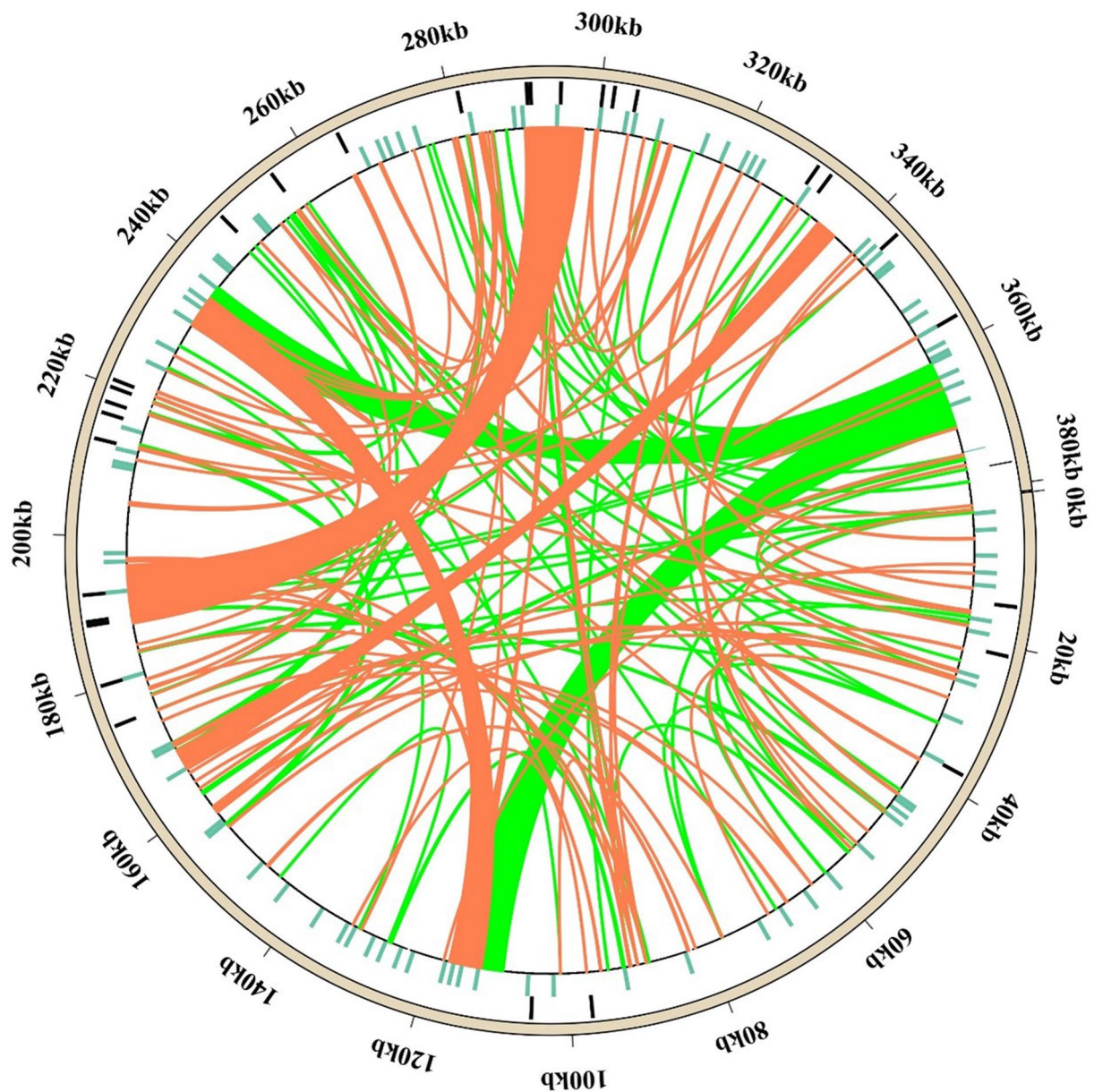
Ou *et al. BMC Plant Biology*   (2024) 24:830

Page 6 of 17

Fig. 2 Distribution of dispersed repeats within the *Agropyron cristatum* mitogenome. The outermost circle represents the mitochondrial genome sequence, followed by tandem repeats, SSR repeats, and dispersed repeats (coral arcs represent 92 forward repeats and the lime represents 93 palindromic repeats

acids were expressed at frequencies exceeding 100. Further analysis showed that all 64 codons in the *A. cristatum* mitogenome were evenly used, with approximately half of them being more widely used (RSCU>1) and half being used less frequently (RSCU<1) (Fig. 5). Among all codons, only the stop codon TGA and tryptophan (TGG) had RSCU values of 1 (Fig. 5), indicating that these two codons were at a normal level. Among the codons with RSCU>1, the most commonly used was the initiation codon methionine (ATG), with an RSCU=3, whereas the termination codon (TAA) had the highest usage, with an RSCU=1.27. Although one amino acid corresponds to

multiple codons, there are varying degrees of preference for codons used in the *A. cristatum* mitogenome. For example, although four codons were used to encode glycine in the *A. cristatum* mitogenome, GGA (RSCU=1.44) and GGT (RSCU=1.29) codons were more frequently used. Moreover, the third base ending in A/T (*n*=28) was more frequently used than the G/C ending (*n*=3) among the preferred codons, which is consistent with the composition of codons in most plant mitogenomes.
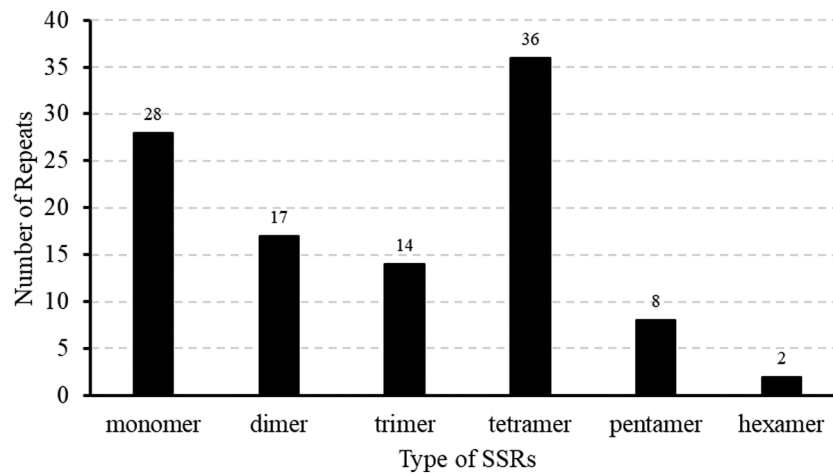
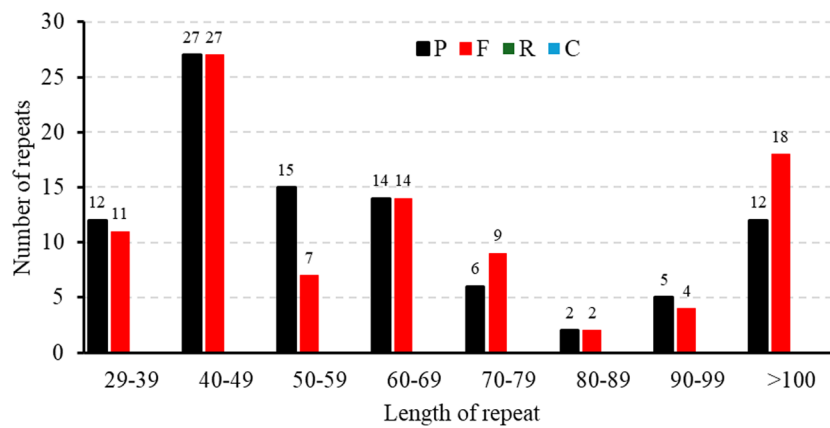**Fig. 3** Distribution of SSRs in the *Agropyron cristatum* mitogenome

Fig. 4 Allocation of the lengths of dispersed repeats in the *Agropyron cristatum* mitogenome. P, palindromic repeat sequences; F, forward repeat sequences; R, Reverse repeat sequences; C, complement repeat sequences

Table 2 Codon counts in the *Agropyron cristatum* mitochondrial PCGs

| Codon	Count	Codon	Count	Codon	Count	Codon	Count
UAA(*)	14	GGC(G)	98	AUG(M)	268	AGU(S)	163
UAG(*)	8	GGG(G)	123	AAC(N)	99	UCA(S)	180
UGA(*)	11	GGU(G)	225	AAU(N)	226	UCC(S)	145
GCA(A)	152	CAC(H)	57	CCA(P)	163	UCG(S)	121
GCC(A)	158	CAU(H)	186	CCC(P)	120	UCU(S)	203
GCG(A)	95	AUA(I)	227	CCG(P)	71	ACA(T)	129
GCU(A)	257	AUC(I)	219	CCU(P)	184	ACC(T)	132
UGC(C)	53	AUU(I)	355	CAA(Q)	221	ACG(T)	74
UGU(C)	91	AAA(K)	260	CAG(Q)	62	ACU(T)	177
GAC(D)	106	AAG(K)	177	AGA(R)	165	GUA(V)	170
GAU(D)	228	CUA(L)	164	AGG(R)	91	GUC(V)	116
GAA(E)	277	CUC(L)	121	CGA(R)	137	GUG(V)	142
GAG(E)	129	CUG(L)	100	CGC(R)	62	GUU(V)	191
UUC(F)	269	CUU(L)	222	CGG(R)	81	UGG(W)	144
UUU(F)	400	UUA(L)	273	CGU(R)	141	UAC(Y)	74
GGA(G)	250	UUG(L)	207	AGC(S)	98	UAU(Y)	239
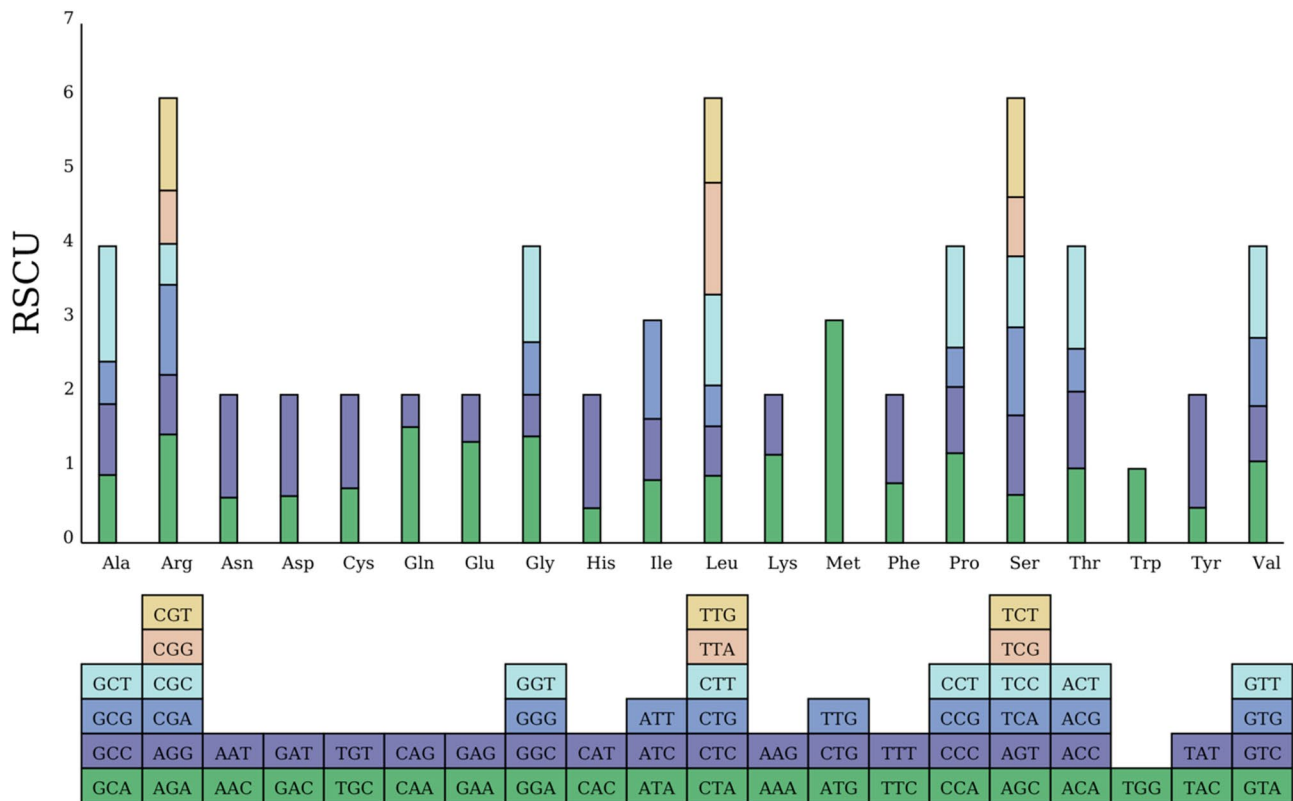
Fig. 5 Relative synonymous codon usage (RSCU) in PCGs of *Agropyron cristatum* mitogenome. Codon families are shown below the graph
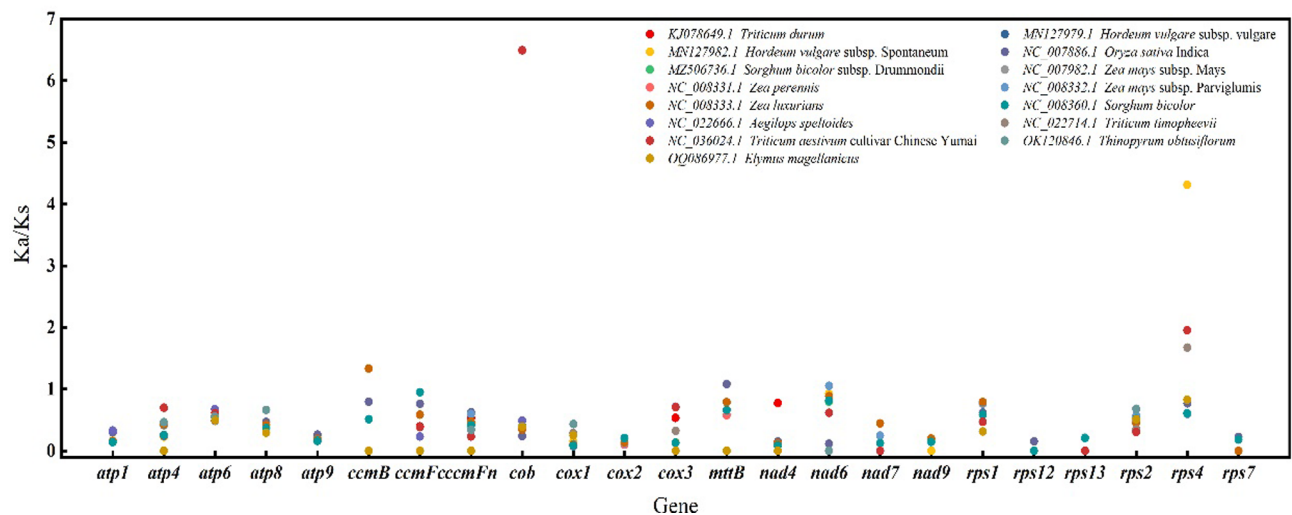

Fig. 6 A dot plot of the Ka/Ks values of 23 protein-coding genes in mitogenomes of *Agropyron cristatum* versus 15 Poaceae species

The substitution rates of mitochondrial PCGs

We selected 15 species of grass plants (including rice, barley, maize, and sorghum) and analyzed the ratio of nonsynonymous-to-synonymous substitutions in 23 PCGs related to mitogenome of *A. cristatum*. The results shown indicated that not all of the 16 species had 23 PCGs (Fig. 6), which were missing to varying degrees in wheat, barley, rice, *Aegilops speltoides*, *Thinopyrum*

obtusiflorum, and *Elymus magellanicus*, whereas were present in sorghum and maize. Among them, *Thinopyrum obtusiflorum* missed 13 PCGs, whereas only two genes were missing in *Oryza sativa subsp. Indica* (Table S7). Of note, within the 23 PCGs of the two species of sorghum, only *rps12* had a $Ka/Ks = Na/0 = 0$ whereas all the remaining genes had $Ka/Ks < 1$, indicating that these genes were more affected by negative selection during
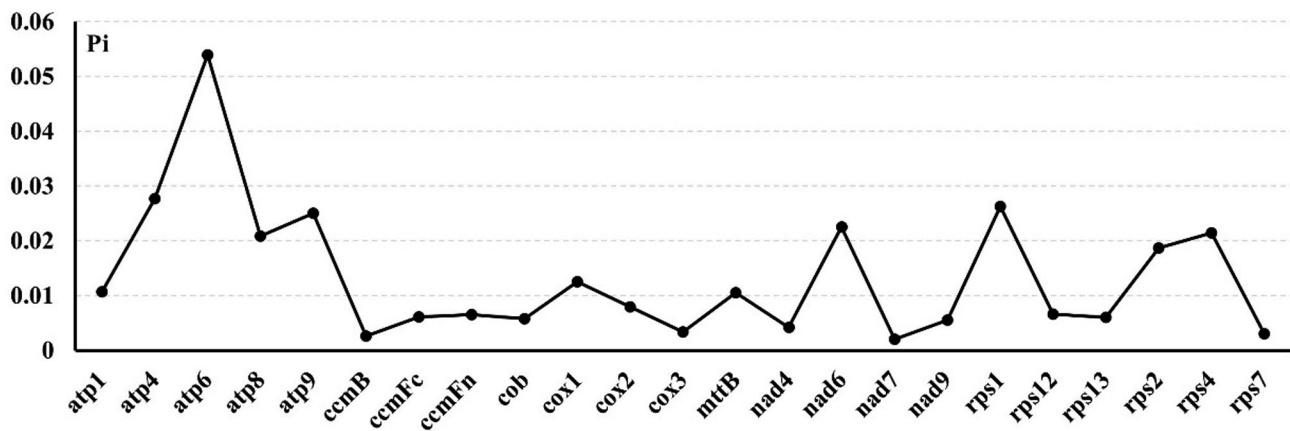
Fig. 7 Nucleotide diversity (Pi) among 23 protein-coding genes in mitogenomes of *Agropyron cristatum*
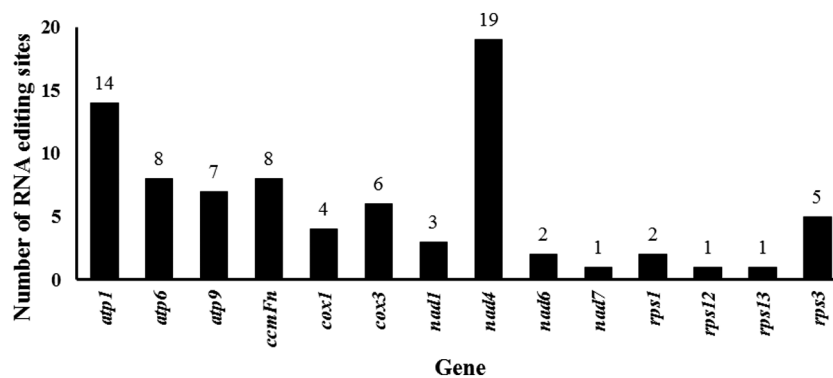

Fig. 8 Distribution of RNA-editing sites in the mitochondrial protein-coding genes of *Agropyron cristatum*

evolution (namely, purifying selection) and showing a conservative tendency towards synonymous substitutions. Among the four species of *Zea*, two genes (*rps12* and *rps7*) had $Ka/Ks=Na/0=0$. In contrast, *ccmB* had $Ka/Ks>1$ in *Zea perennis* and *Zea luxurians* (both 1.33467), and *nad6* had $Ka/Ks>1$ in *Zea mays subsp. Mays* and *Zea mays subsp. Parviglumis* (both 1.0524). These results indicated that although most genes in the 15 plant species showed a more conservative evolutionary trend, there were still individual genes with very high positive selection effects. For example, *rps4* had a $Ka/Ks>1.5$ in six plants species, including *Triticum*, *Hordeum vulgare*, and *Aegilops*, with a $Ka/Ks=4.30778$ in two *Hordeum* spp., and *cob* had a $Ka/Ks=6.48797$ in two wheat plants (*Triticum durum* and *Triticum aestivum*). These results indicate that the genes in these plants underwent dramatic and important changes during their evolution, which are of great significance for their own development.

Nucleotide diversity

To further explore the selection variation of genes under different selective pressures in the mitochondria of *Agropyron cristatum*, we conducted nucleotide diversity on 23

PCGs in the substitution rates. Among these 23 PCGs, the range of Pi values is from 0.0021 to 0.0539, with an average of 0.013. Among them, the gene *atp6* has the highest variability (Pi=0.0539), while *nad7* (Pi=0.0021) and *ccmb* (Pi=0.0025) have the lowest, with a huge gap of 25 times between them. At the same time, we also noticed that nearly half of the 23 protein-coding genes (*n*=11) have a polymorphism greater than 0.01, with a much larger variation range than those genes with polymorphism less than 0.01. From the perspective of functional enrichment, these 23 PCGs show higher variability in genes related to ATP synthase function, while genes involved in a series of activities such as Cytochrome c have lower performance.

RNA editing sites prediction

RNA editing in plants primarily occurs in organelles such as mitochondria and chloroplasts, and is closely related to the functions of these organelles. Therefore, for the study of the mitogenome in *A. cristatum*, RNA editing is a very important molecular mechanism [15, 39]. Through the analysis and prediction of PCGs in the mitogenome of *A. cristatum*, we identified 81 effective RNA editing sites distributed across 14 PCGs (Fig. 8). Of note, *nad4*

and *atp1* were the genes with most RNA editing sites (19 and 14 sites, respectively), whereas *nad7*, *rps12*, and *rps13* only had one RNA editing site. The number of change sites in the remaining genes ranged from three to eight (Table S8). There were 11 types of changes in the identified 81 RNA editing sites, mainly C to T changes (*n*=53, 65.4%). In addition, except for the mutual changes in A and G that occurred more than five times, the frequency of the other change types was less than four. The RNA editing sites identified involve changes in 56 codons, which mainly at the second base (*n*=43, 53.1%), followed by the first base (*n*=23, 28.4%), with the least changes occurring at the third base (*n*=15, 18.5%). During the RNA editing process, 39 types of amino acid changes were involved, including eight synonymous and two special (to stop codons) changes. Among these, the most common was the conversion of serine to leucine (*n*=11). The PCGs involved in RNA-editing site events in the mitogenome of *A. cristatum* accounted for only 40% of the total 35 PCGs, with an average of 5.8 predicted change sites per gene. Amino acid variations mainly affected leucine, arginine, serine, and threonine (55.6%). Taken together, these results indicate that the scale and extent of RNA editing of the mitogenome of *A. cristatum* are not very large and that the overall trend is more for it to be conservative and stable.

Chloroplast-derived mitogenomic sequences

As semi-autonomous organelles, mitochondria and chloroplasts share many similarities and even identical genes, and both exist within the cell in an endosymbiotic manner. Frequent DNA transfer has been reported to occur between the mitochondria, chloroplasts, and nuclear genomes during the evolution [10, 13, 14]. Therefore, we compared the mitochondrial and chloroplast genomes of *A. cristatum* to have a better view of this flow of DNA fragments (Fig. 9). Overall, the chloroplast genome (135,554 bp) [3] of *A. cristatum* is much smaller than its mitochondrial genome (381,065 bp) and a total of 95 transferred fragments were identified. These fragments comprised the vast majority of rRNA and tRNA genes, as well as a small portion of the IGS region of PCGs. The total length of the transferred fragments was 34,343 bp, covering 9% of the mitogenome. More over almost no perfectly matched or complete fragments were within these transferred fragments, and the maximum matching degree that could be maintained was 97.826%, which was achieved by the transference of *trnN-GUU* from the chloroplasts to the mitochondria. These DNA fragments underwent diverse transfer phenomena between genes, gene spacers, introns, and gene spacers, indicating that the gene exchange between the two organelle genomes was repetitive, random, and rearranged. Throughout these processes, the longest fragments transferred were

rps7, *rps12* (intron), *rps12*, *ndhB* (intron), and *ndhB* transferred from the chloroplast to the mitochondrial IGS (*nad4L*, *nad7*), with a total length of 4,357 bp and a complete matching degree of (98.531%). The most frequently transferred fragments were *rrn16* and *rrn23* from the chloroplast to *rrn18* and *rrn26* from the mitochondria (24 and 20 times, respectively). Meanwhile, our results showed that in the case of all mitochondrial transfers and maintenance of the integrity of most fragments, there were two rRNA genes (*rrn18* and *rrn26*), nine tRNA genes (*trnM-CAU*, *trnF-GAA*, *trnC-GCA*, *trnS-GGA*, *trnS-UGA*, *trnN-GUU*, *trnW-CCA*, *trnP-UGG*, *trnQ-UUG*), and two PCGs (*nad5* and *rps12*). During sequence migration between the mitochondrial and chloroplast genomes of *A. cristatum*, the tRNA genes exhibited more stable evolutionary characteristics.

Phylogenetic analysis

We selected 20 species of grasses, including *A. cristatum*, and two species of legumes, and integrated their published mitogenomes to construct a phylogenetic tree (Fig. 10) to further explore the mitogenome evolution pattern within a large population. Overall, the 22 species were divided into three groups, one of which was composed of *Triticum*, *Hordeum*, *Aegilops*, *Elymus*, and *Agropyron*; the second was composed of rice, sorghum, and corn; and the third group was composed of soybeans and alfalfa (outgroups). The ML support values were in good agreement with those obtained for these groups. More than 12 nodes in the 19 nodes of the phylogenetic tree had support rates above 90%, including nine nodes with 100% support, indicating that these results were reliable. The tree clearly divided legumes and grasses, and distinguished plants between different genera. *A. cristatum* was in the first group with the largest number of classified species with close genetic relationship with *Triticum*, *Thinopyrum*, and *Aegilops* as compared with *Elymus* and *Hordeum*. In addition, there were more complex trends in the evolutionary classification of the four wheat species and two *Aegilops* species, which could only be preliminarily classified using mitogenomes. Further exploration will require additional features to refine the results. Nonetheless, this analysis proves mitogenomes can be effectively used for the phylogenetic analysis of species and exploring their genetic relationships.

Discussion

Structural characteristics of the mitogenome in *agropyron cristatum*

The genomes of organelles such as mitochondria and plastids, which exist in an endosymbiotic manner and have independent heritability, have been gradually explored with the advancement of genetic analysis methods [15, 44]. Research on eukaryotic mitogenomes
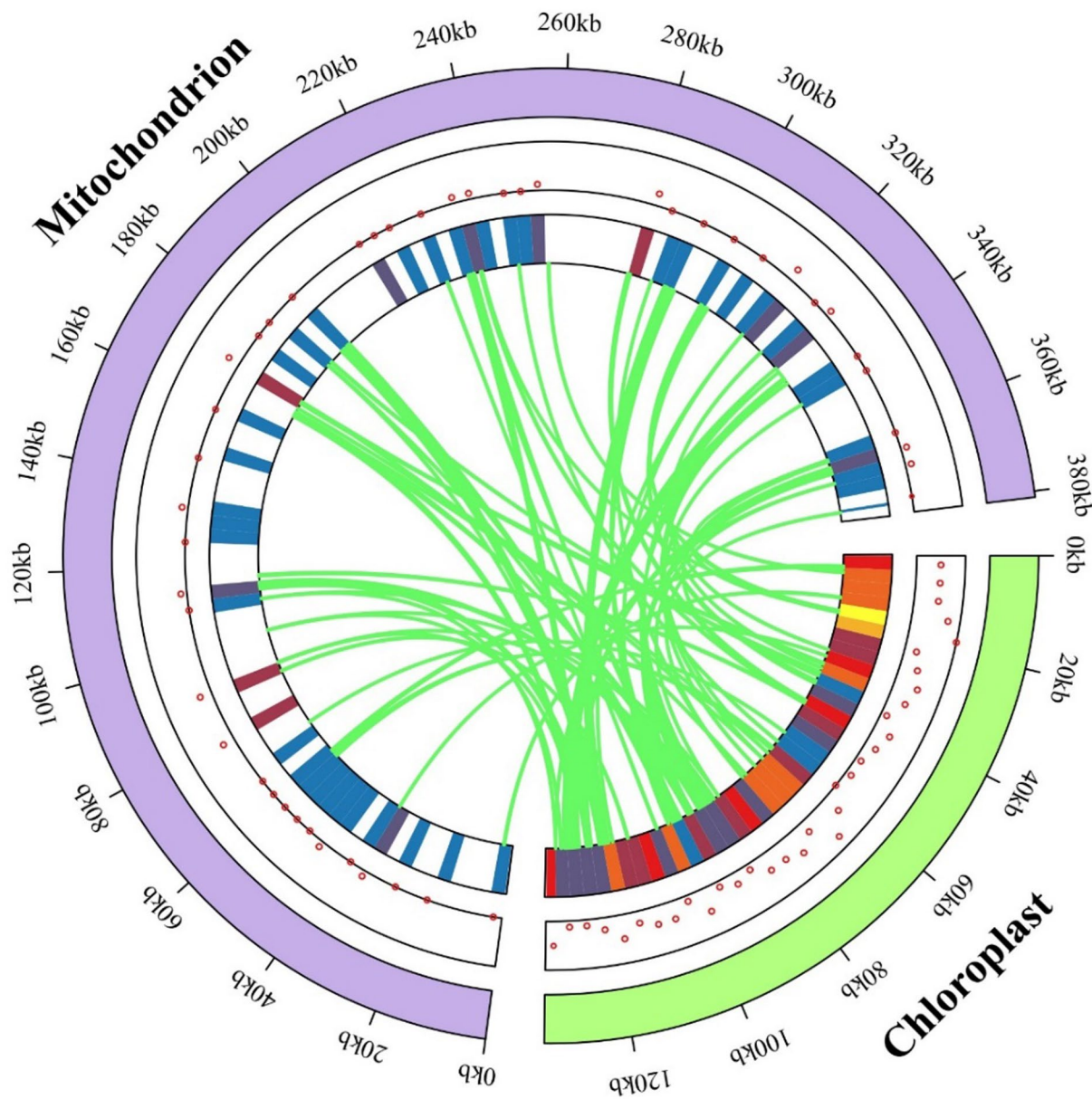
Fig. 9 Gene transfer events between the chloroplast and mitochondrial genome. Dots and heatmaps inside the two chromosomes demonstrate where the migrated genes are located. The green and purple circular outer segments represent the chloroplast and mitochondrial genome. The green lines inside the circle portray the migration routes of chloroplast-like sequences identified in the mitogenome

has focused on animals, whereas research on plants has focused on the unique chloroplast genomes [45]. Using bioinformatics technology, we constructed a single circular molecule of 381,065 bp in length for the *A. cristatum* mitogenome to complement the missing link in the understanding of *A. cristatum* genetic information. To date, the largest mitogenome described in plants was larger than 10 Mb (*L. sibirica*) [12], whereas the smallest was only 6.6 kb and belonged to the genus *Plasmodium* [46]. It can be seen that the size of the entire plant mitochondrial genome is in the middle and upper reaches of eukaryotic organisms, while the size of the

A. cristatum mitochondrial genome in this study is not prominent compared to other Poaceae plants (Table S1, >400 kb). This indicates that the *A. cristatum* mitochondrial genome has fewer introns, repetitive sequences, and sequence transfers, and may be in a more conserved state during the evolution of grass plants. The composition of the.

A. cristatum mitogenome structure was consistent with the usual circular molecular structure of plant mitochondria [7, 47, 48]. Nonetheless, additional findings into the mitogenome structure have suggested that, due to limitations in prior assumptions, assembly methods,
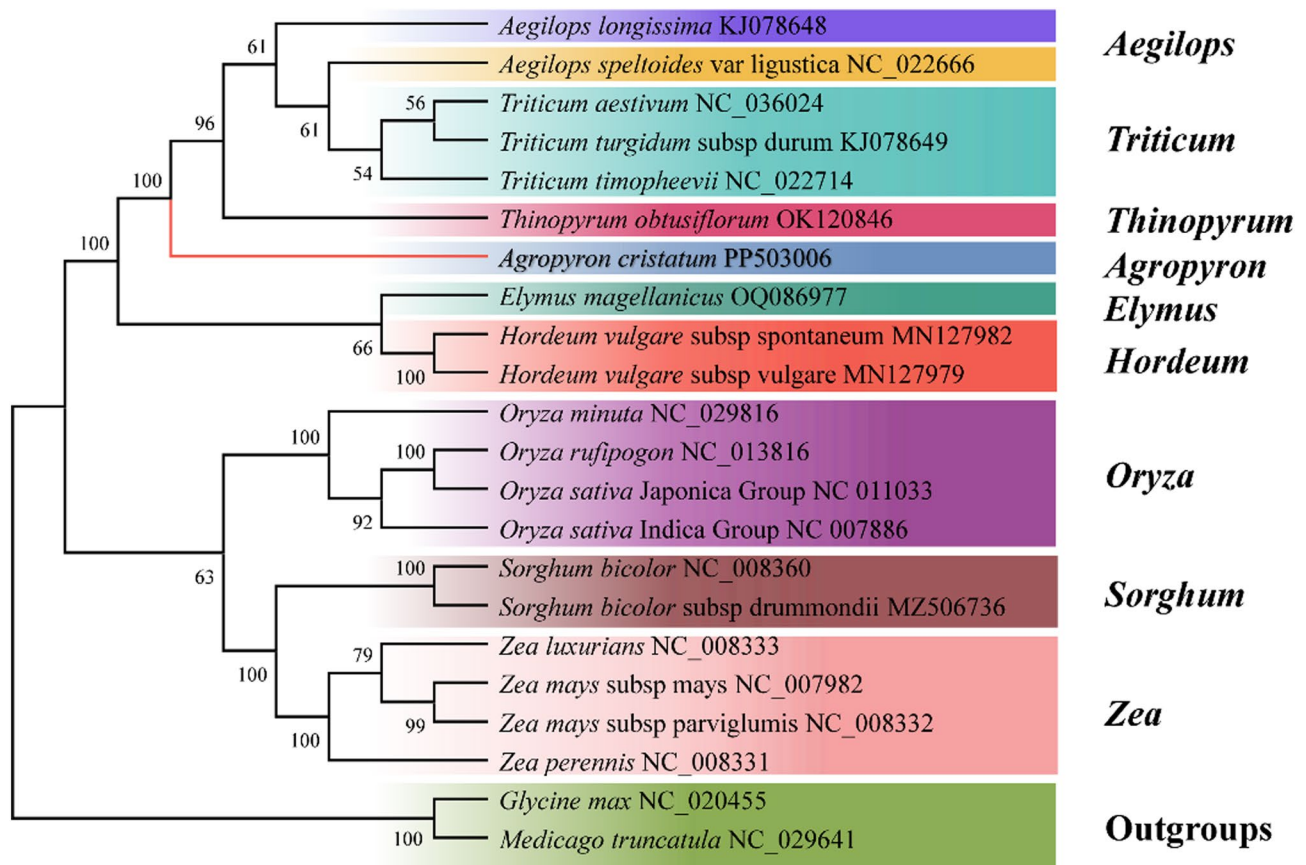
Fig. 10 Representation of the phylogenetic relationships of *Agropyron cristatum* with 19 Poaceae spp. The Maximum Likelihood bootstrap values are shown for each node

and parameter constraints, the structure of mitochondria can be more accurately identified as a dynamic, multiform collection [8, 49–52]. Therefore, research on the configuration of the *A. cristatum* mitogenome should be conducted by considering the frequent changes in gene sequences, large fragment repetitive sequences, and intuitive image illustrations to better explore its morphology and variation characteristics.

Although plant mitogenomes exhibit greater lengths and more complex variations than animal mitogenomes [7, 12, 13], their coding sequences remain highly conserved [53–55]. Previous studies have identified 41 PCGs (24 core and 17 variable genes) in angiosperms [7, 55], whereas we annotated 14 core and 21 variable genes (including two pseudogenes) in *A. cristatum* mitogenome. The core genes are relatively low in number, but the variable genes exhibit high abundance. These differences were due to the loss and transfer of PCGs in the mitogenome during evolution. For example, almost all PCGs in the *Zostera* mitogenome were lost [55], whereas the largest number of PCGs ($n=42$) was reported in the liverwort *Haplomitrium* [56, 57]. Moreover, significant differences were observed in the number of mitochondrial PCGs among different plants. Of note, the nucleus,

plastid, and mitochondria genomes frequently undergo DNA fragment transfer during evolution, but complete loss of gene function during major genetic changes can lead to the deletion of one or more PCGs. Therefore, based on the number of mitochondrial PCGs annotated in *A. cristatum*, we preliminarily determined that there were drastic genetic changes in *Agropyron* plants during evolution from a common ancestor to better adapt to their environment.

Diversity and variation in the mitogenome of *agropyron cristatum*

Many repetitive sequences in the *A. cristatum* mitogenome we identified, including simple sequence, tandem, and dispersed repeats. These repetitive sequences provide sufficient basic conditions for molecular recombination in the mitochondria of different sizes and numbers [58]. At the same time, the specific genetic information they contain can often be used as an important genetic analysis tool [59, 60]. Previous studies have shown that the size of the mitogenome is positively correlated with the size of repetitive sequences that cover more than 500 bp [61, 62]. We identified a total of 320 repetitive sequences in the *A. cristatum* mitogenome, with a total length of

381,065 bp and covering more than 10% of the mitogenome. Similarly, the repetitive sequences in watermelon cover 10% (379,236 bp), whereas those in *Cucurbita pepo* (982,833 bp) have tens of thousands of repetitive sequences, covering 38% of the entire mitogenome [61]. These results further support the relationship between repetitive sequences size and the size of the mitogenome. In addition, ultralong sequences that appear in these repetitive sequences are of interest, as they are more likely to interact with structural changes in genes and thus play a driving role in the evolution of higher plants. In *G. barbadense*, ultralong repetitive sequences result in the duplication of seven genes [59]. Similarly, ultralong repetitive sequences in the mitochondria of *Arabidopsis thaliana* (6.5 and 6.2 kb) and sugar beet (6.2 kb) were reported to play a significant role in promoting molecular recombination [63, 64]. In the wheatgrass mitochondrial genome, five repetitive sequences exceeding 1,000 bp resulted in the duplication of six genes (*rrn26, rrn18, rrn5, ccmC, trnP-UGG, trnM-CAU*). Therefore, repetitive sequences present in the mitogenome are an important source of molecular rearrangements in the mitochondria.

RNA editing can also promote specific and important changes in the expression pattern of genes in the mitochondria [65]. We found 81 RNA editing sites in the mitogenome of *A. cristatum*, which is far fewer than that those in rice (*n*=491) [66] and *Arabidopsis thaliana* (441) and *Brassica napus* (*n*=427) [65] which all below to the same family (Gramineae). The main type of editing changes that occurred at *A. cristatum* RNA editing sites was a transition from C to T, which is consistent with the general type of RNA editing change in plants [67]. The second type of editing change mainly occurred at the first and second base positions, which is also consistent with most of the editing characteristics previously reported [68, 69]. This result may be due to the degeneracy of codons, which is not retained because of synonymous changes, or it may also be due to the lack of efficiency of the analysis methods used [60, 65, 70]. In addition, in the mitogenome of *A. cristatum*, RNA editing mainly affected 14 PCGs comprised within five functional groups: ATP synthase, Cytochrome c oxidase, Cytochrome c biogenesis, NADH dehydrogenase, and Ribosomal proteins (SSU). In most plants, rRNA and tRNA genes almost never undergo RNA editing [71]. Moreover, we identified more ATP synthase- and NADH dehydrogenase-related genes being affected by RNA editing as compared with two species of *Phaseolus vulgaris* [70] and *Acer truncatum Bunge* [60], which cytochrome c biogenesis- and NADH dehydrogenase-related genes were the most affected. It can be hypothesized that this different may be related to the functional bias and environmental adaptation of the species.

Gene flow between the nuclear and mitochondrial genomes of plant cells is bidirectional, which provides additional variation in the evolution of plant genomes and makes their evolutionary relationships more complex [72]. It is generally believed that the direction of DNA fragment transfer in plants is mainly from the plastid genome to the mitogenome; therefore, different levels of DNA fragments from the chloroplast can be identified in the mitogenome of plants [73]. For example, 0.56%, 10.85%, and 16.34% of plastid-derived sequences were identified in the mitogenomes of *Marchantia polymorpha* [74], *Phoenix dactylifera* [75], and *R. chingii* [76], respectively. We can roughly determine that the range of sequence coverage in plant mitochondria transferred from plastids is 0–20%. In our study, we found that the total length of chloroplast-derived fragments in the mitogenome of *A. cristatum* is 34,343 bp, covering 9% of the mitochondrial genome, which suggest frequent DNA fragment flow between the mitochondria and chloroplasts. During the transfer of chloroplast genes to mitochondrial genes, there are both complete functional retention and changes in gene structure after transfer, as well as different types of gene transfer from gene spacers to complete genes. We also found that protein-coding and rRNA genes usually undergo changes in gene structure after transfer, whereas tRNA genes retain the greatest degree of gene functional integrity during the transfer process. Based on this, it can be inferred that the tRNA genes may share certain characteristics between the two organelles (*trnM-CAU, trnF-GAA, trnC-GCA, trnS-GGA, trnS-UGA, trnN-GUU, trnW-CCA, trnP-UGG, trnQ-UUG*), which are also reflected in the DNA transfer process between the two organelles in soybeans [77].

Evolutionary classification of *agropyron cristatum* based on mitogenome

The preference for codon usage is not only a characteristic of species evolution but also a way to distinguish between nuclear and mitochondrial genomes [78]. The changes in synonymous codons that occur during this process do not affect the final protein changes and are therefore considered selection neutral [78–80]. However, both synonymous and non-synonymous codon changes lead to preferential usage of PCGs [78, 81]. The phenomenon of codon usage preference reflects a bias in base composition and the trend of natural selection within the genome, to some extent [82, 83]. In the mitogenome of *A. cristatum*, the codon with the highest preference for usage was the initiation codon (methionine) ATG, followed by (glutamine) CAA, (alanine) GCT, (histidine) CAT, (tyrosine) TAT, which are similar to the frequently used codons in *Forsythia suspensa* [69] and *Quercus acutissima* [84]. Moreover, we also found that the number of codons with preference changes ending with A/T at the

third base was much higher than of those ending with G/C, which occurs in most species [85, 86], indicating the preference of *A. cristatum* mitochondrial genes for base changes.

Ka/Ks ratio is another important indicator of the degree of sequence homology variation and evolutionary relationships between PCGs in different species [87]. Calculations of the ratio of non-synonymous-to-synonymous substitutions can help characterize the positive, neutral, or negative selection status of the PCG in different species during the evolutionary process [69, 87]. Using *A. cristatum* as reference, we analyzed 23 PCGs in 15 species of *Gramineae* and the results showed that more than 80% of the genes had *Ka/Ks* values below 1, which suggested that ATP synthase-, cytochrome c oxidase-, NADH dehydrogenase-related genes were more affected by purifying selection during the evolutionary process were in a conservative state as compared with *A. cristatum*. However, *rps4*, *cob*, *mttB*, and *ccmB* were positively selected in different species. Synonymous substitutions are more commonly reported in plants and the variation bias of these important genes is limited by purifying selection [88]. Through ratio analysis of nonsynonymous and synonymous substitution rates, we can observe subtle changes in homologous PCGs in different species, which can further predict the evolutionary trends of species during evolution and support the construction of more accurate evolutionary systems and taxonomies.

To validate the classification of plants in the *Agropyron* genus using mitogenomes, we constructed a phylogenetic tree containing 22 plant mitogenomes. Previously, the genus *Agropyron* was defined as containing plants, such as *Thinopyrum* and *Elymus*, in a broad sense. Later the classification of plants in the *Agropyron* genus included species within the crown *Agropyron* complex containing the P genome, including *A.cristatum*, *A.desertorun*, *A.michnoi*, *A.sibiricum*, *A.mongolicum* and others [89]. *Agropyron* are polyploid plants that are homologous polyploids or segmental allopolyploids, and their intraspecific genetic complexity is certain. Morphological classification and genetic analysis have confirmed that the P genome of the *Agropyron* genus is not closely related to that of other genera in the *Triticeae* family and has a high degree of independent heritability [5]. Similarly, in the constructed tree, plants of the genera *Thinopyrum*, *Elymus*, *Triticum*, and *Agropyron* were in the same large group, and *Agropyron* was in a separate branch, which is perfect agreement with previous studies. This also indicates that classification based on mitogenomes is applicable to the classification of plants, such as *Agropyron*, and further supports the classification status of *Agropyron* spp. in the *Triticeae* family of *Poaceae*, while providing a reference for further exploration of the intraspecific phylogenetic relationships of *Agropyron* plants. In summary, although we have conducted an analysis of phylogenetics and genetic evolution based on the characteristics of the mitogenome of *A. cristatum*, we have not yet delved deeply into the detailed evolutionary history and biogeography of its closely related species. Future work could further refine and explore these topics to provide a more comprehensive analysis.

Conclusions

In this study, we assembled the first mitogenome structure of *Agropyron* and conducted preliminary characterization. The mitogenome of *A. cristatum* is a single circular molecular structure, with a full length of 381,065 bp, comprising 52 genes (35 protein-coding, 3 rRNA, and 14 tRNA genes). Next, we analyzed the diverse changes in DNA fragments, including three types of repetitive sequences, RNA editing site prediction, and gene transfer patterns between organelles, to clarify the three sources and modes of gene conformational changes in the mitochondria of *A. cristatum*. Finally, using codon bias, non-synonymous and synonymous substitution rates, nucleotide diversity, and phylogenetic methods, we confirmed the independent genetic status and general evolutionary trends of *Agropyron* plants in the *Triticeae* family. In summary, this study deepens our understanding of the phylogeny and genetic evolution of the *Agropyron* genus from the perspective of the mitogenome, addressing a gap in genomic research on this genus. This lays a solid theoretical foundation and provides valuable scientific basis for future genetic breeding research and for the development and utilization of the valuable genetic features of *Agropyron*.

Abbreviations
rRNA	Ribosomal RNA
tRNA	Transfer RNA
PCGs	Protein-coding genes
IGS	Intergenic spacer
SSRs	Simple sequence repeats
RSCU	Relative synonymous codon usage ratios
ML	Maximum Likelihood

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12870-024-05558-8.

Supplementary Material 1. Additional file: Table S1. GenBank accession numbers of plant mitogenomes sampled in this study. Table S2. Size of introns and exons in split genes in *Agropyron cristatum* mitogenome. Table S3. Microsatellite repeats in *Agropyron cristatum* mitogenome. Table S4. Distribution of SSRs in *Agropyron cristatum* mitogenome. Table S5. Dispersed repeats in *Agropyron cristatum* mitogenome. Table S6. Distribution of tandem repeats in in *Agropyron cristatum* mitogenome. Table S7. The pairwise Ka/Ks values among the 15 Poaceae species. Table S8. RNA editing sites in the mitogenome of *Agropyron cristatum*. Table S9. Fragments transferred from chloroplast to mitochondria in *Agropyron cristatum*.

Ou *et al. BMC Plant Biology*   (2024) 24:830

Page 15 of 17

Declarations

Ethical approval and consent to participate
All materials were collected from Hohhot, Inner Mongolia, China (40.57°N, 111.93°E) by us and deposited in the National Medium-Term Genebank Forage Germplasm (accession number: CF000396). It was identified by Dr. Zinian Wu of Institute of Grassland Research, Chinese Academy of Agricultural Sciences, Hohhot, China. The study strictly adhered to all relevant institutional, national and international guidelines and legislation. We confirm that no additional specific permits were required for the collection of plant samples.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Institute of Grassland Research, Chinese Academy of Agricultural Sciences, Hohhot, China
[2]Key Laboratory of Grassland Resources and Utilization of Ministry of Agriculture, Hohhot, China

References
1. Wang RRC. *Agropyron* and *Psathyrostachys*. Wild Crop Relatives: Genomic Breed. Resour. 2011;77–108.
2. LI HY, LI YX, LI J, WU ZN, HUANG F, ZHU L, GUO MW, LI ZY. XIN X. Phenotypic Diversity Analysis and comprehensive evaluation of 143 *Agropyron* Germplasm resources in Inner Mongolia. J Plant Genetic Resour.2024;1–15.
3. Chen N, Sha LN, Dong ZZ, Tang C, Wang Y, Kang HY, Zhang HQ, Yan XB, Zhou YH, Fan X. Complete structure and variation of the chloroplast genome of *Agropyron cristatum* (L.) Gaertn. Gene. 2018;640:86–96.
4. Zhou S, Zhang J, Che Y, Liu W, Lu Y, Yang X, Li X, Jia J, Liu X, Li L. Construction of *Agropyron* Gaertn. Genetic linkage maps using a wheat 660K SNP array reveals a homoeologous relationship with the wheat genome. Plant Biotechnol J. 2018;16(3):818–27.
5. Dewey DR. The genomic system of classification as a guide to intergeneric hybridization with the perennial *Triticeae*/. In: Gustafson JP, editor. Gene Manipulation in Plant Improvement. New York: Plenum; 1984. pp. 209–79.
6. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. Sequence and organization of the human mitochondrial genome. Nature. 1981;290(5806):457–65.
7. Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, Christensen AC. The alternative reality of plant mitochondrial DNA: one ring does not rule them all. PLoS Genet. 2019;15(8):e1008373.
8. Sloan DB. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. New Phytol. 2013;200(4):978–85.
9. Oldenburg DJ, Bendich AJ. DNA maintenance in plastids and mitochondria of plants. Front Plant Sci. 2015;6:883.
10. Guo W, Grewe F, Fan W, Young GJ, Knoop V, Palmer JD, Mower JP. Ginkgo and Welwitschia Mitogenomes reveal Extreme contrasts in Gymnosperm mitochondrial evolution. Mol Biol Evol. 2016;33(6):1448–60.
11. Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant *Viscum Scurruloideum* is extremely divergent and dynamic and has lost all nad genes. Proc Natl Acad Sci U S A. 2015;112(27):E3515–24.
12. Putintseva YA, Bondar EI, Simonov EP, Sharov VV, Oreshkova NV, Kuzmin DA, Konstantinov YM, Shmakov VN, Belkov VI, Sadovsky MG, Keech O, Krutovsky KV. Siberian larch (*Larix sibirica* Ledeb.) Mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. BMC Genomics. 2020;21(1):654.
13. Bergthorsson U, Adams KL, Thomason B, Palmer JD. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature. 2003;424(6945):197–201.
14. Wynn EL, Christensen AC. Repeats of unusual size in Plant mitochondrial genomes: identification, incidence and evolution. (Bethesda). 2019;G3(2):549–59.
15. Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. The plant mitochondrial genome: dynamics and maintenance. Biochimie. 2014;100:107–20.
16. Wang N, Li C, Kuang L, Wu X, Xie K, Zhu A, Xu Q, Larkin RM, Zhou Y, Deng X, Guo W. Pan-mitogenomics reveals the genetic basis of cytonuclear conflicts in citrus hybridization, domestication, and diversification. Proc Natl Acad Sci USA. 2022;119(43):e2206076119.
17. Dewey DR. .Historical and current taxonomic perspectives of *Agropyron*, *Elymus*, and related Genera1.Crop science. 1983;23(4).
18. Linc G, Gaál E, Molnár I, Icsó D, Badaeva E, Molnár-Láng M. Molecular cytogenetic (FISH) and genome analysis of diploid wheatgrasses and their phylogenetic relationship. PLoS ONE. 2017;12(3):e0173623.
19. Che Y, Yang Y, Yang X. Phylogenetic relationship and diversity among *Agropyron Gaertn*. Germplasm using SSRs markers. Plant Syst Evol. 2015;301:163–70.
20. Chen SY, Ma X, Zhang XQ, Huang LK, Zhou JN. Genetic diversity and relationships among accessions of five crested wheatgrass species (Poaceae: *Agropyron*) based on gliadin analysis. Genet Mol Res. 2013;12(4):5704–13.
21. Zhou S, Yan B, Li F, Zhang J, Zhang J, Ma H, Liu W, Lu Y, Yang X, Li X, Liu X, Li L. RNA-Seq analysis provides the First insights into the phylogenetic relationship and interspecific variation between *Agropyron cristatum* and wheat. Front Plant Sci. 2017;8:1644.
22. Han H, Qiu R, Liu Y, Zhou X, Gao C, Pang Y, Zhao Y. Analysis of Chloroplast genomes provides insights into the evolution of *Agropyron*. Front Genet. 2022;13:832809.
23. Sahlin K, Baudeau T, Cazaux B, Marchet C. A survey of mapping algorithms in the long-reads era. Genome Biol. 2023;24(1):133.
24. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722–36.
25. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
26. Wick RR, Judd LM, Gorrie CL, Holt KE, Unicycler. Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595.
27. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015;31(20):3350–2.
28. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–11.
29. Yu JJ, Yang WB, Song JB, dePamphilis Y, Yi CW, Li TS DZ, editors. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;10;21(1):241.
30. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 2019;15:50.
31. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47(W1):W59–64.
32. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5.

33. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
34. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29(22):4633–42.
35. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017;34(12):3299–302.
36. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
37. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–95.
38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9. 1000 Genome Project Data Processing Subgroup.
39. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.
40. Wu S, Liu W, Aljohi HA, Alromaih SA, Alanazi IO, Lin Q, Yu J, Hu S. REDO: RNA editing detection in Plant Organelles based on variant calling results. J Comput Biol. 2018;25(5):509–16.
41. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative Toolkit developed for interactive analyses of big Biological Data. Mol Plant. 2020;13(8):1194–202.
42. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9.
43. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22(21):2688–90.
44. Mower JP. Variation in protein gene and intron content among land plant mitogenomes. Mitochondrion. 2020;53:203–13.
45. Friedman JR, Nunnari J. Mitochondrial form and function. Nature. 2014;505(7483):335–43.
46. Hikosaka K, Watanabe Y, Tsuji N, Kita K, Kishine H, Arisue N, Palacpac NM, Kawazu S, Sawai H, Horii T, Igarashi I, Tanabe K. Divergence of the mitochondrial genome structure in the apicomplexan parasites, Babesia and Theileria. Mol Biol Evol. 2010;27(5):1107–16.
47. Lonsdale DM, Hodge TP, Fauron CM. The physical map and organisation of the mitochondrial genome from the fertile cytoplasm of maize. Nucleic Acids Res. 1984;12(24):9249–61.
48. Palmer JD, Shields CR. Tripartite structure of the *Brassica campestris* mitochondrial genome. Nature. 1984;307(5950):437–40.
49. Oldenburg DJ, Bendich AJ. Size and structure of replicating mitochondrial DNA in cultured Tobacco cells. Plant Cell. 1996;8(3):447–61.
50. Bendich AJ. Structural analysis of mitochondrial DNA molecules from fungi and plants using moving pictures and pulsed-field gel electrophoresis. J Mol Biol. 1996;255(4):564–88.
51. Mower JP, Case AL, Floro ER, Willis JH. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. Genome Biol Evol. 2012;4(5):670–86.
52. Oldenburg DJ, Bendich AJ. Mitochondrial DNA from the liverwort *Marchantia polymorpha*: circularly permuted linear molecules, head-to-tail concatemers, and a 5' protein. J Mol Biol. 2001;310(3):549–62.
53. Adams KL, Qiu YL, Stoutemyer M, Palmer JD. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci U S A. 2002;99(15):9905–12.
54. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 2003;29(3):380–95.
55. Petersen G, Cuenca A, Zervas A, Ross GT, Graham SW, Barrett CF, Davis JI, Seberg O. Mitochondrial genome evolution in Alismatales: size reduction and extensive loss of ribosomal protein genes. PLoS ONE. 2017;12(5):e0177606.
56. Dong S, Zhao C, Zhang S, Zhang L, Wu H, Liu H, Zhu R, Jia Y, Goffinet B, Liu Y. Mitochondrial genomes of the early land plant lineage liverworts (*Marchantiophyta*): conserved genome structure, and ongoing low frequency recombination. BMC Genomics. 2019;20(1):953.
57. Myszczyński K, Ślipiko M, Sawicki J. Potential of Transcript Editing across Mitogenomes of Early Land Plants Shows Novel and Familiar trends. Int J Mol Sci. 2019;20(12):2963.
58. Alverson AJ, Zhuo S, Rice DW, Sloan DB, Palmer JD. The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats. PLoS ONE. 2011;6(1):e16404.
59. Bi C, Paterson AH, Wang X, Xu Y, Wu D, Qu Y, Jiang A, Ye Q, Ye N. Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium Raimondii* by Comparative Genomics approaches. Biomed Res Int. 2016;2016:5040598.
60. Ma Q, Wang Y, Li S, Wen J, Zhu L, Yan K, Du Y, Ren J, Li S, Chen Z, Bi C, Li Q. Assembly and comparative analysis of the first complete mitochondrial genome of *Acer Truncatum Bunge*: a woody oil-tree species producing nervonic acid. BMC Plant Biol. 2022;22(1):29.
61. Allen JO, Fauron CM, Minx P, Roark L, Oddiraju S, Lin GN, Meyer L, Sun H, Kim K, Wang C, Du F, Xu D, Gibson M, Cifrese J, Clifton SW, Newton KJ. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics. 2007;177(2):1173–92.
62. Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol. 2010;27(6):1436–48.
63. Unseld M, Marienfeld JR, Brandt P, Brennicke A. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat Genet. 1997;15(1):57–61.
64. Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA(Cys)(GCA). Nucleic Acids Res. 2000;28(13):2571–6.
65. Handa H. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and Arabidopsis thaliana. Nucleic Acids Res. 2003;31(20):5907–16.
66. Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol Genet Genomics. 2002;268(4):434–45.
67. Small ID, Schallenberg-Rüdinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. Plant organellar RNA editing: what 30 years of research has revealed. Plant J. 2020;101(5):1040–56.
68. Sloan DB, MacQueen AH, Alverson AJ, Palmer JD, Taylor DR. Extensive loss of RNA editing sites in rapidly evolving Silene mitochondrial genomes: selection vs. retroprocessing as the driving force. Genetics. 2010;185(4):1369–80.
69. Song Y, Du X, Li A, Fan A, He L, Sun Z, Niu Y, Qiao Y. Assembly and analysis of the complete mitochondrial genome of *Forsythia suspensa* (Thunb.) Vahl. BMC Genomics. 2023;24(1):708.
70. Bi C, Lu N, Xu Y, He C, Lu Z. Characterization and analysis of the mitochondrial genome of Common Bean (*Phaseolus vulgaris*) by comparative genomic approaches. Int J Mol Sci. 2020;21(11):3778.
71. Giegé P, Brennicke A. RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. Proc Natl Acad Sci U S A. 1999;96(26):15324–9.
72. Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, dePamphilis CW, Knox EB, Palmer JD. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm Amborella. Science. 2013;342(6165):1468–73.
73. Lai C, Wang J, Kan S, Zhang S, Li P, Reeve WG, Wu Z, Zhang Y. Comparative analysis of mitochondrial genomes of *Broussonetia spp.* (Moraceae) reveals heterogeneity in structure, synteny, intercellular gene transfer, and RNA editing. Front Plant Sci. 2022;13:1052151.
74. Zhao N, Grover CE, Chen Z, Wendel JF, Hua J. Intergenomic gene transfer in diploid and allopolyploid *Gossypium*. BMC Plant Biol. 2019;19(1):492.
75. Zhao N, Wang Y, Hua J. The roles of Mitochondrion in Intergenomic Gene Transfer in plants: a source and a Pool. Int J Mol Sci. 2018;19(2):547.
76. Sun M, Zhang M, Chen X, Liu Y, Liu B, Li J, Wang R, Zhao K, Wu J. Rearrangement and domestication as drivers of *Rosaceae* mitogenome plasticity. BMC Biol. 2022;20(1):181.
77. Chang S, Wang Y, Lu J, Gai J, Li J, Chu P, Guan R, Zhao T. The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels. PLoS ONE. 2013;8(2):e56502.
78. Hugaboom M, Hatmaker EA, LaBella AL, Rokas A. Evolution and codon usage bias of mitochondrial and nuclear genomes in *Aspergillus section Flavi*. G3 (Bethesda). 2023;13(1):jkac285.
79. Jia W, Higgs PG. Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection. Mol Biol Evol. 2008;25(2):339–51.

80. LaBella AL, Opulente DA, Steenwyk JL, Hittinger CT, Rokas A. Variation and selection on codon usage bias across an entire subphylum. PLoS Genet. 2019;15(7):e1008304.
81. Hanson G, Coller J. Codon optimality, bias and usage in translation and mRNA decay. Nat Rev Mol Cell Biol. 2018;19(1):20–30.
82. Wei L, He J, Jia X, Qi Q, Liang Z, Zheng H, Ping Y, Liu S, Sun J. Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. BMC Evol Biol. 2014;14:262.
83. Sharp PM, Stenico M, Peden JF, Lloyd AT. Codon usage: mutational bias, translational selection, or both? Biochem Soc Trans. 1993;21(4):835–41.
84. Liu D, Guo H, Zhu J, Qu K, Chen Y, Guo Y, Ding P, Yang H, Xu T, Jing Q, Han S, Li W, Tong B. Complex physical structure of complete mitochondrial genome of *Quercus acutissima* (Fagaceae): a significant energy plant. Genes (Basel). 2022;13(8):1321.
85. Ke SJ, Liu DK, Tu XD, He X, Zhang MM, Zhu MJ, Zhang DY, Zhang CL, Lan SR, Liu ZJ. *Apostasia* Mitochondrial Genome Analysis and *Monocot* Mitochondria Phylogenomics. Int J Mol Sci. 2023;24(9):7837.
86. Guo Y, Li Z, Jin S, Chen S, Li F, Wu H. Assembly and comparative analysis of the Complete Mitochondrial Genome of Two Species of Calla Lilies (*Zantedeschia*, *Araceae*). Int J Mol Sci. 2023;24(11):9566.
87. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteom Bioinf. 2010;8(1):77–80.
88. Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol Biol. 2007;7:135.
89. Dewey DR. .Historical and current taxonomic perspectives of Agropyron, Elymus, and related Genera1.Crop science. 1983;23(4).

Publisher's note