

Additional Methods

K-mer analysis

To gain preliminary insights into the mitochondrial genomes, k-mer analysis was performed using KMC3 v3.2.1 [1]. Firstly, 17-mers were counted and converted into frequency using `kmc_dump` in the KMC3 package. Then the k-mer frequency was visualized using a custom R script (Fig. S1). Mitochondrial genome sizes for each species were evaluated using KmerGenie v1.7051 [2].

Assembly error correction

Clean sequencing reads were mapped back to the assembled genome sequences using Burrows-Wheeler alignment tool BWA v0.7.17 [3]. Assembly errors were identified by inspecting those improper-mapping reads, which included the reads with unexpected insert sizes ($>500\text{bp}$), conflicted read orientations (RF, RR, FF) and long-soft clipping ($\geq 50\text{bp}$). Those scaffolds were further joined using reference sequences based on *de bruijn* graph (NC_007579 for *Triticum aestivum*, KJ078649 for *Triticum turgidum*, and KJ078648 for *Aegilops tauschii*). Sequence regions with low kmer depth were finally confirmed using Integrative Genomics Viewer v2.14.0 [4] with BAM files generated by BWA.

The mitochondrial sequences (mtDNAs; 452,522bp) in *Triticum aestivum* (AABBDD) and *T. turgidum* assembled in this study were slightly shorter than previously reported sequences: NC_007579 (452,528 bp, Sanger assembly) [5], NC_036024 (452,526 bp, Sanger assembly) [6], and MH051716 (452,526bp, NGS 200x) [7]. Manual curations verified that the differences were most caused by single base insertions or deletions in low-complexity intergenic regions, which did not affect the genic sequences. Since our sequencing data were much deeper than those assemblies, we were confident to proceed.

Sequencing library evaluation

Clean sequencing reads were mapped back to the final curated genome sequences using BWA v0.7.17 [3]. Duplicated reads were further removed using Picard v2.25.0

(<https://broadinstitute.github.io/picard/>). The fragment sizes were calculated based on the starting and ending position for each read pair using the Python module pysam, and then were plot using a custom R script (Fig. S3). Sequencing depth was calculated using bamCoverage program in DeepTools v3.5.1 [8] at a bin size of 50bp, and then plotted using a custom R script.

Domain and motif analysis

In order to compare the domains and motifs among 4 most variable genes, the known protein domains were analyzed using InterProScan (v5.61; Database v93.0) [9] with the parameters -iprlookup --minsize 6 (Pfam v35.0). The new motifs were discovered using the MEME Suite v5.4.1 [10] with parameters -protein -evt 10.0 -nmotifs 10 -minw 6 -maxw 300 -mod anr. Then these domains and motifs were illustrated using a custom R script based on the drawProteins v1.18.0 package [11].

Sequence similarity searches against nuclear genomes

Gene sequences were searched against nuclear genomes using local blastn in BLAST+ v2.13.0 [12] with parameters: -eval 1e-6 -outfmt '6 qseqid sseqid pident length mismatch gapopen qstart qend sstart send eval bitscore qlen slen' -perc_identity 95. The latest versions of genomic sequences were downloaded from NCBI/Ensembl/IWGSC: *Hordeum vulgare* (HH) MorexV3 (GCA_904849725.1), *Triticum urartu* (AA) Tu2.1 (GCA_003073215.2), *Aegilops tauschii* (DD) Aet5.0 (GCA_002575655.2), *Ae. speltoides* (SS) ASM2143724v1 (GCA_021437245.1), *T. turgidum* (AABB) Svevo1 (GCA_900231445.1), and *T. aestivum* IWGSC2.1 (AABBDD) GCA_018294505.1.

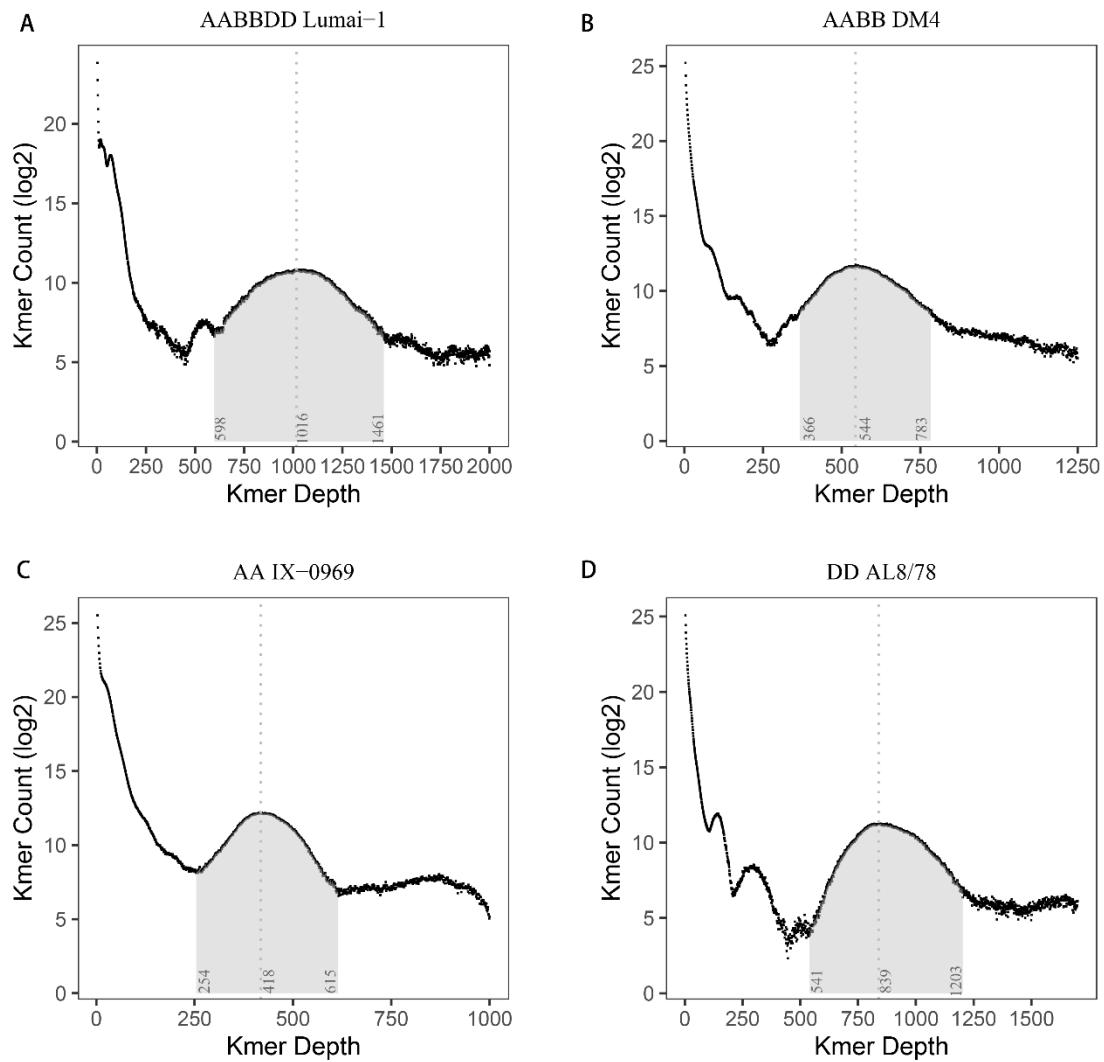


Fig. S1. Kmer frequency for genome size estimation. A) Hexaploid wheat *Triticum aestivum* (AABBDD) Lumai-1; B) Tetraploid wheat *T. turgidum* (AABB) DM4; C) Diploid *T. urartu* (AA) IX-0969; D) Diploid *Aegilops tauschii* (DD) AL8/78.

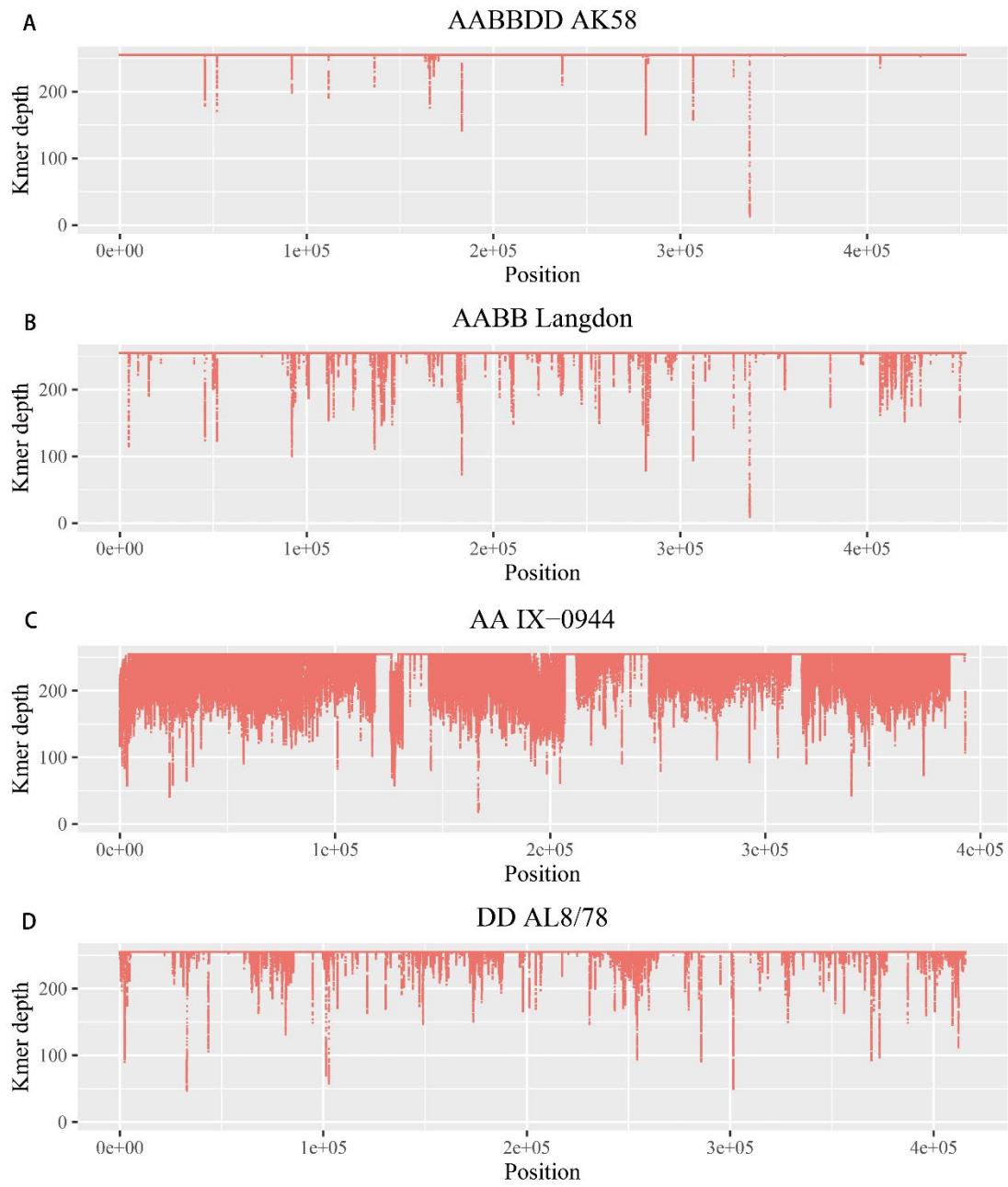


Fig. S2. Detection of assembly error check by k-mer depth. Larger kmer depth were set to 255 for easy visualization of **A)** *Triticum aestivum* (AABBDD) AK58, **B)** *T. turgidum* (AABB) Langdon, **C)** *T. urartu* (AA) IX-0944, and **D)** *Aegilops tauschii* (DD) AL8/78.

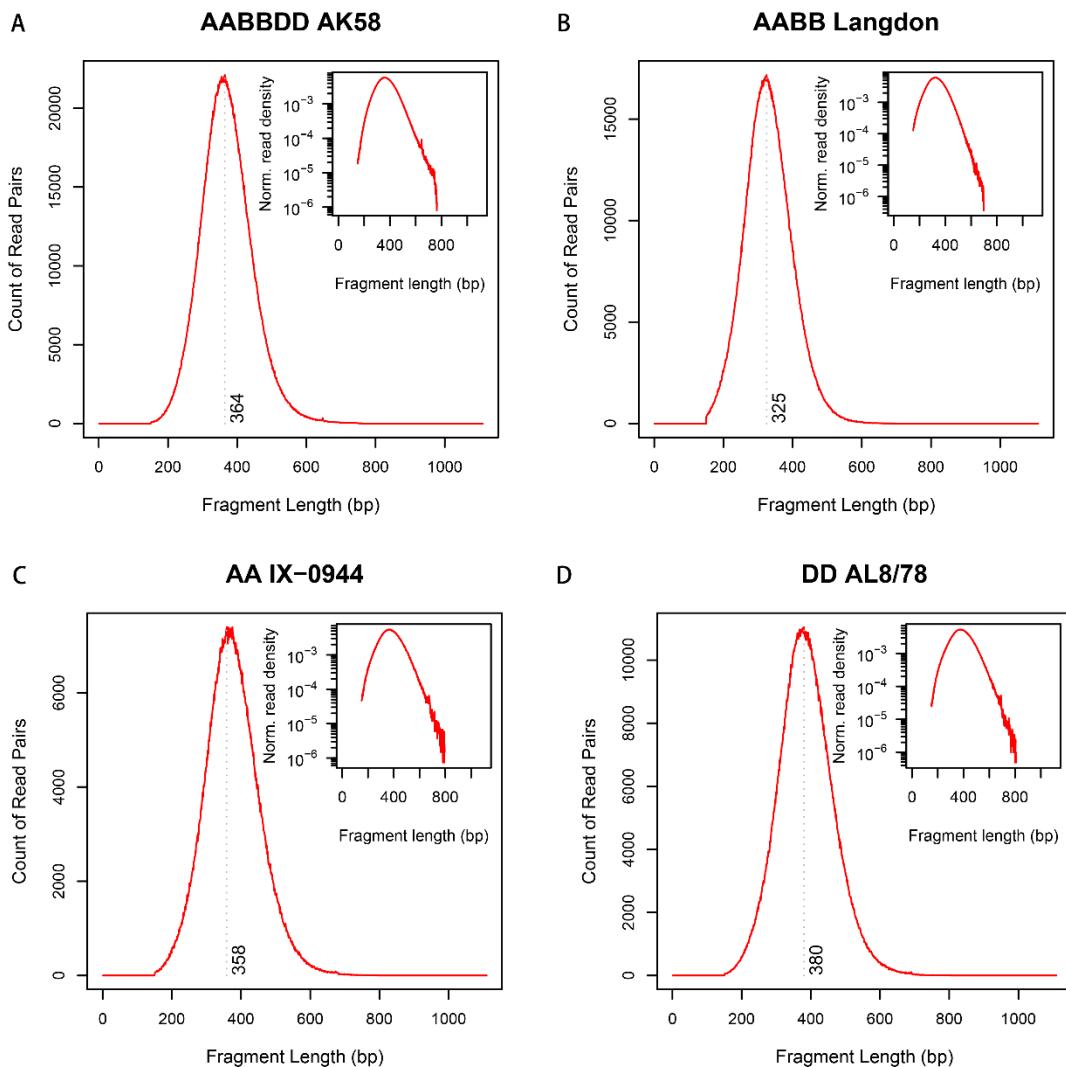


Fig. S3. Fragment size distribution of representative sequencing libraries. **A)** *Triticum aestivum* (AABBDD) AK58, **B)** *T. turgidum* (AABB) Langdon, **C)** *T. urartu* (AA) IX-0944, and **D)** *Aegilops tauschii* (DD) AL8/78.

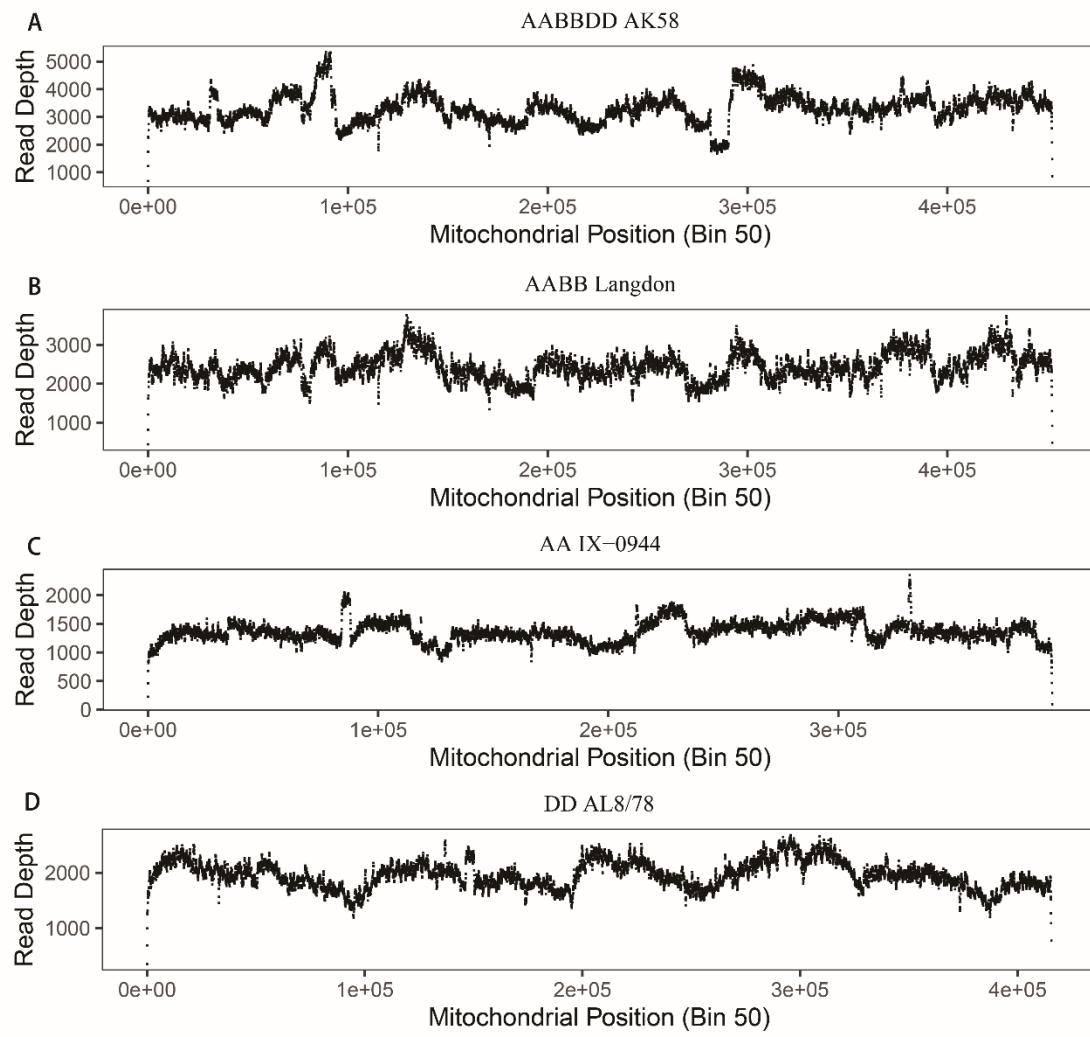


Fig. S4. Sequencing depth of unique reads. The duplicated reads were removed to minimize the errors brought by sequencing library construction. **A)** *Triticum aestivum* (AABBDD) AK58, **B)** *T. turgidum* (AABB) Langdon, **C)** *T. urartu* (AA) IX-0944, and **D)** *Aegilops tauschii* (DD) AL8/78.

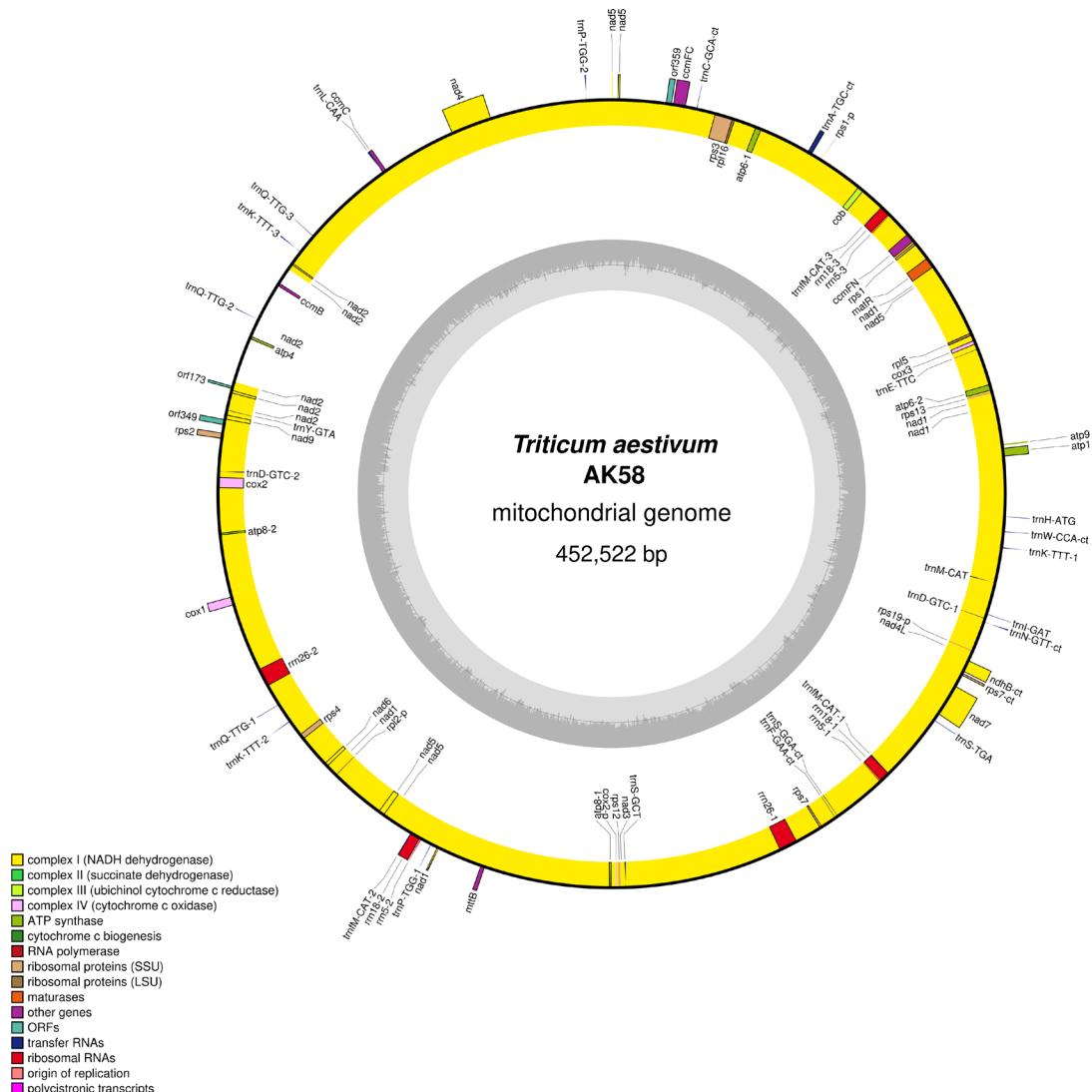


Fig. S5. Sequence map of the mitochondrial genome in *Triticum aestivum* AK58

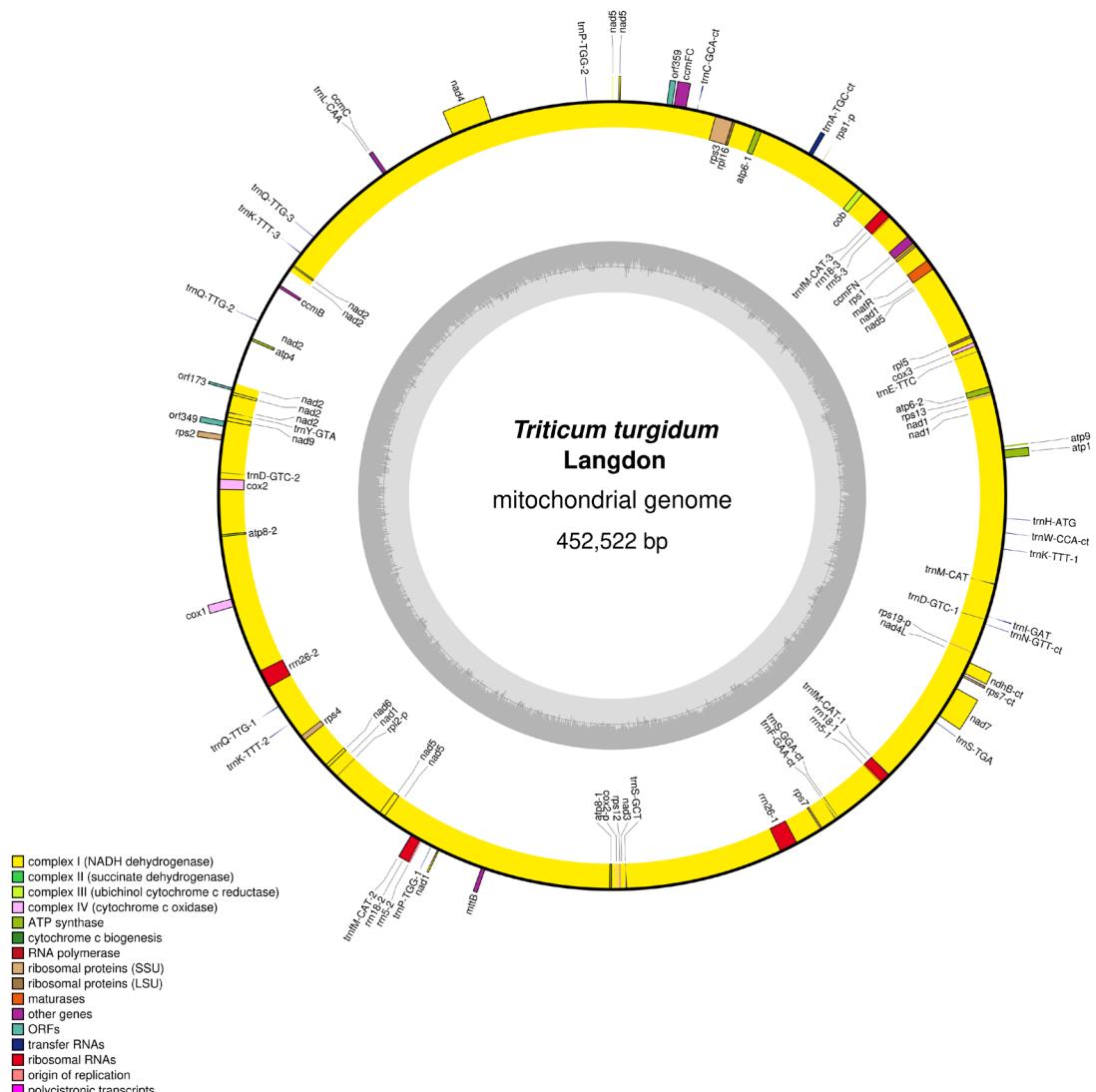


Fig. S6. Sequence map of the mitochondrial genome in *Triticum turgidum* Langdon

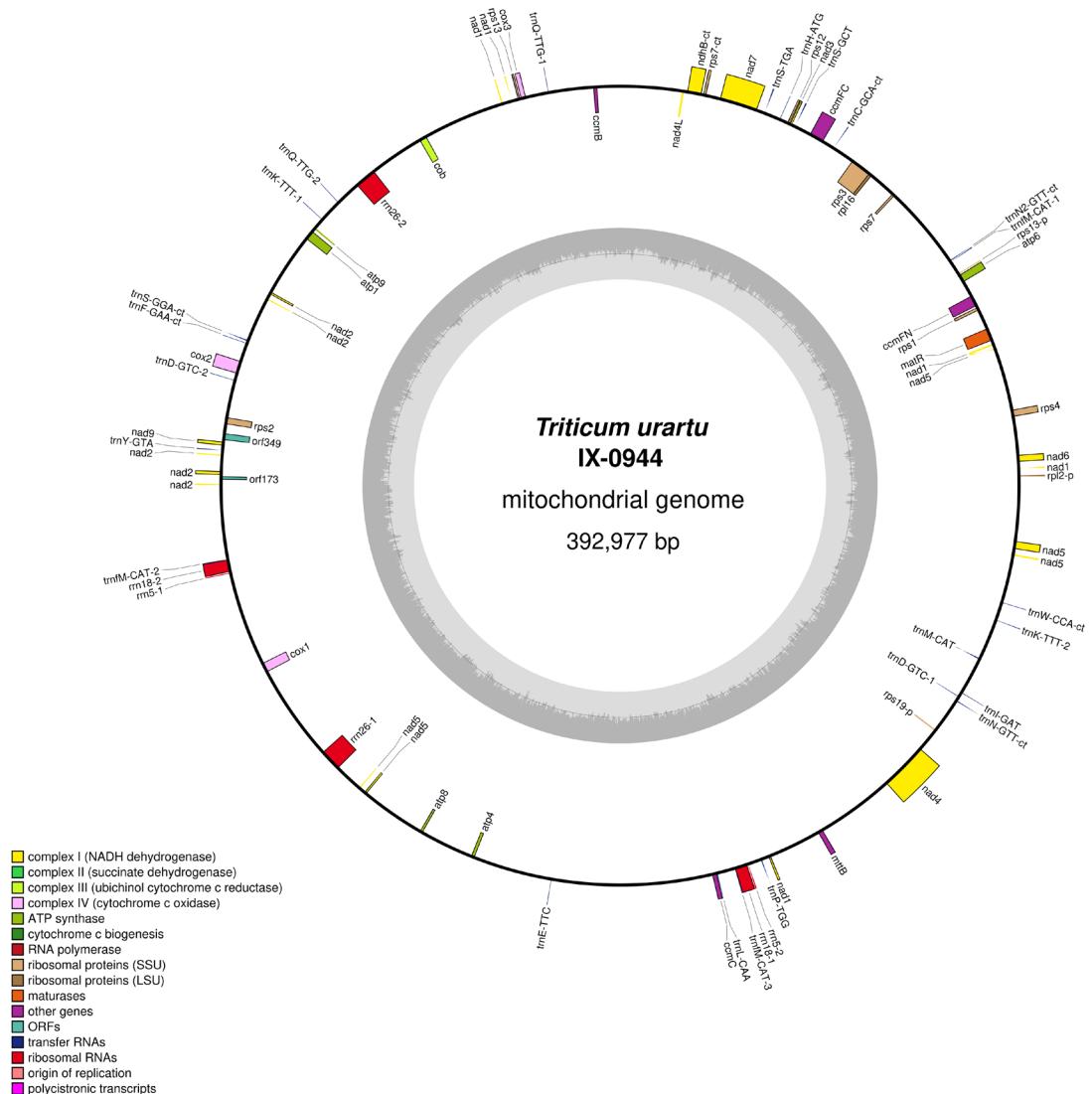


Fig. S7. Sequence map of the mitochondrial genome in *Triticum urartu* IX-0944

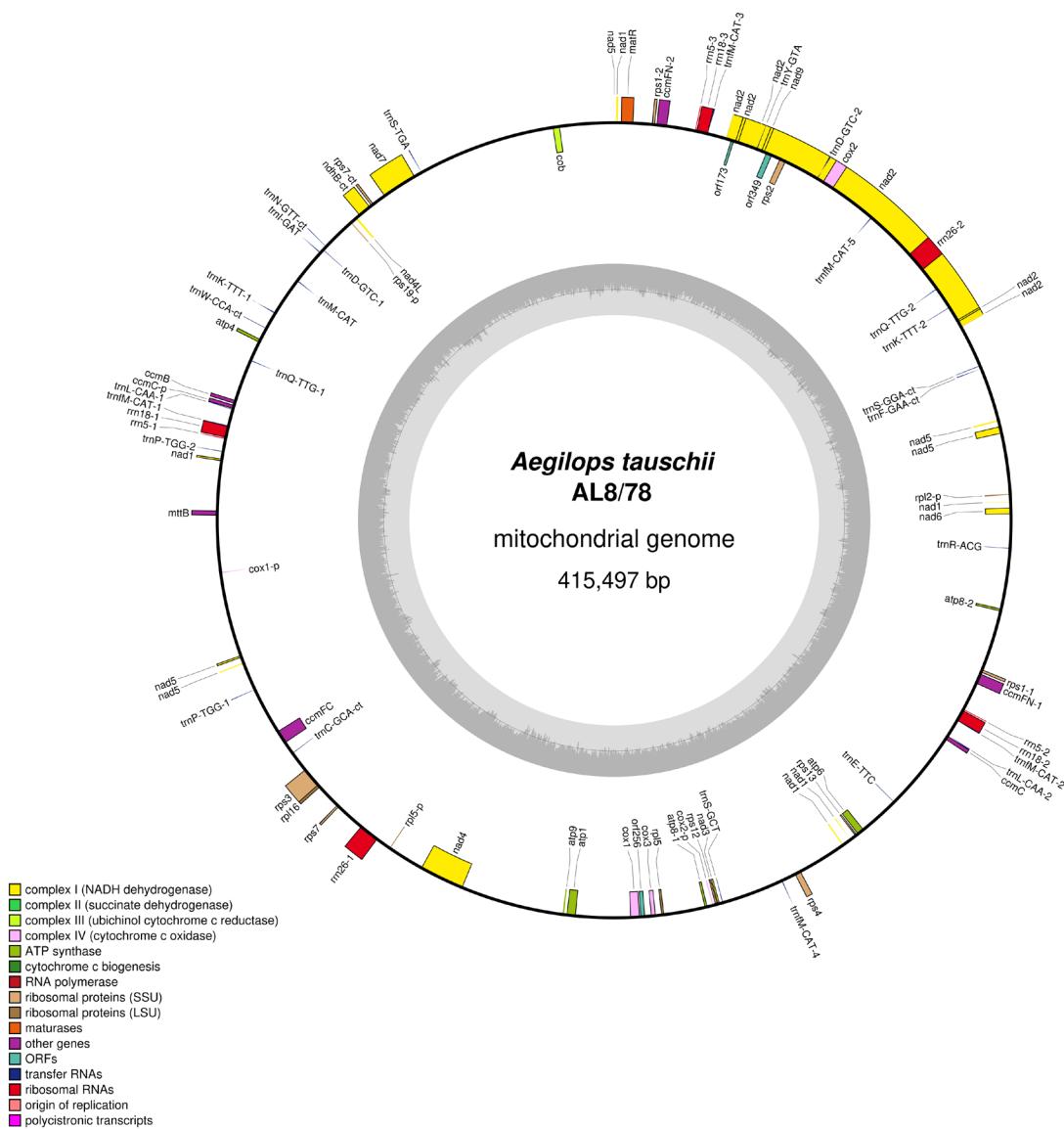


Fig. S8. Sequence map of the mitochondrial genome in *Aegilops tauschii* AL8/78

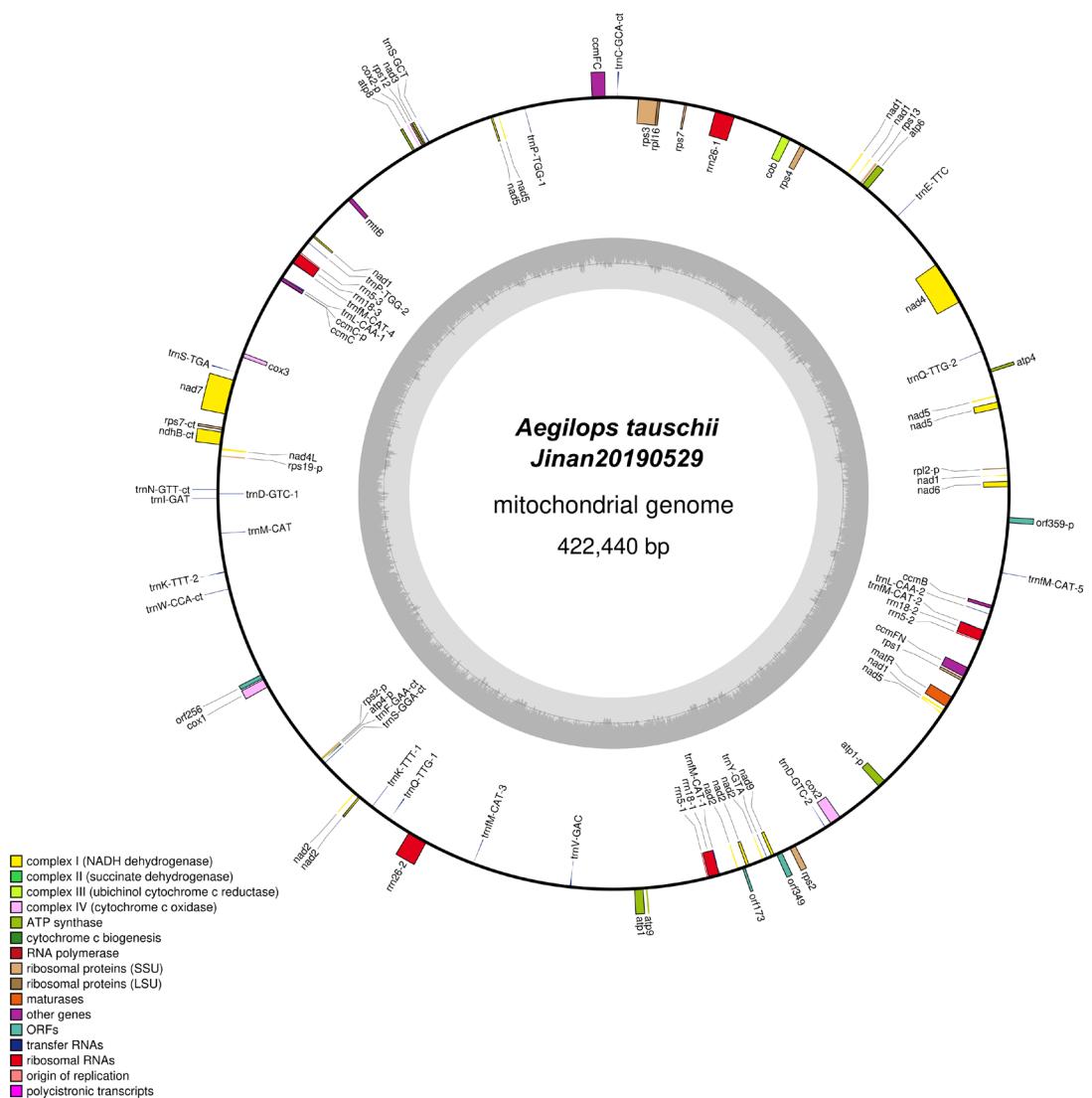


Fig. S9. Sequence map of the mitochondrial genome in *Aegilops cylindrica* Jinan190529

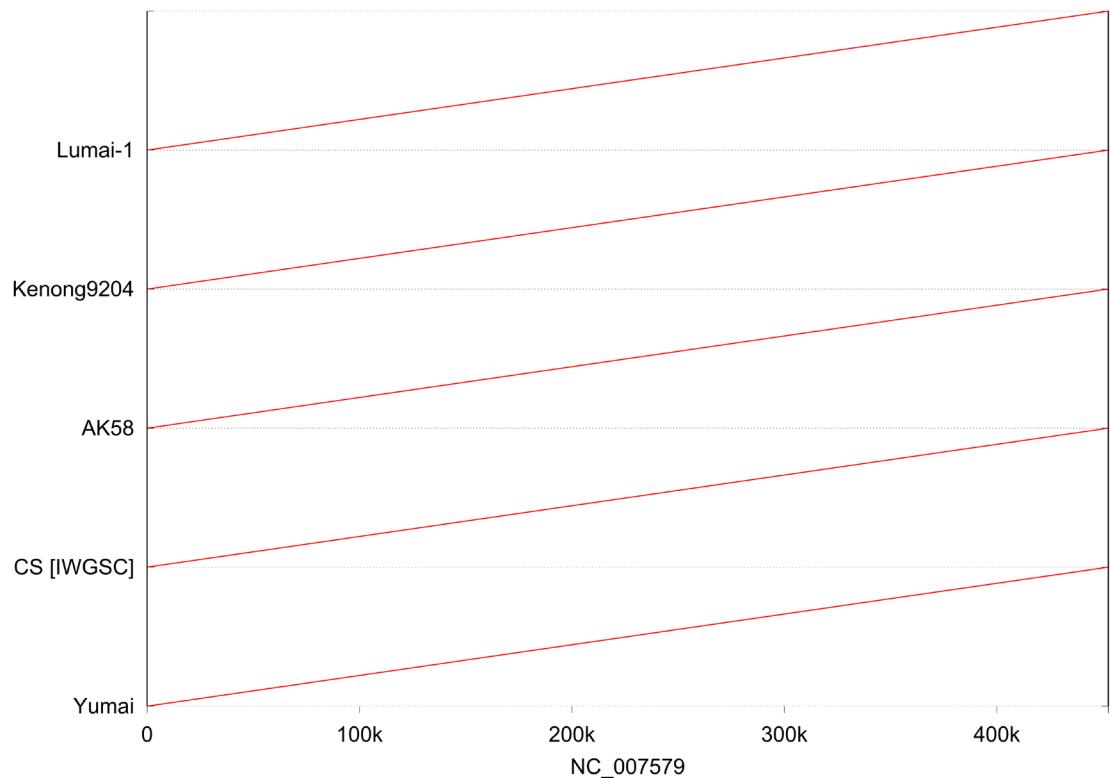


Fig. S10. **MUMmerplot of assembled mitochondrial genomes in hexaploid wheats.** *Triticum aestivum* Lumai-1, *T. aestivum* Kenong9204, *T. aestivum* AK58; CS[IWGSC]: *T. Chinese Spring* 42, MH051716; Yumai: *T. aestivum* Yumai NC_036024; and *T. aestivum* Chinese Spring NC_007579

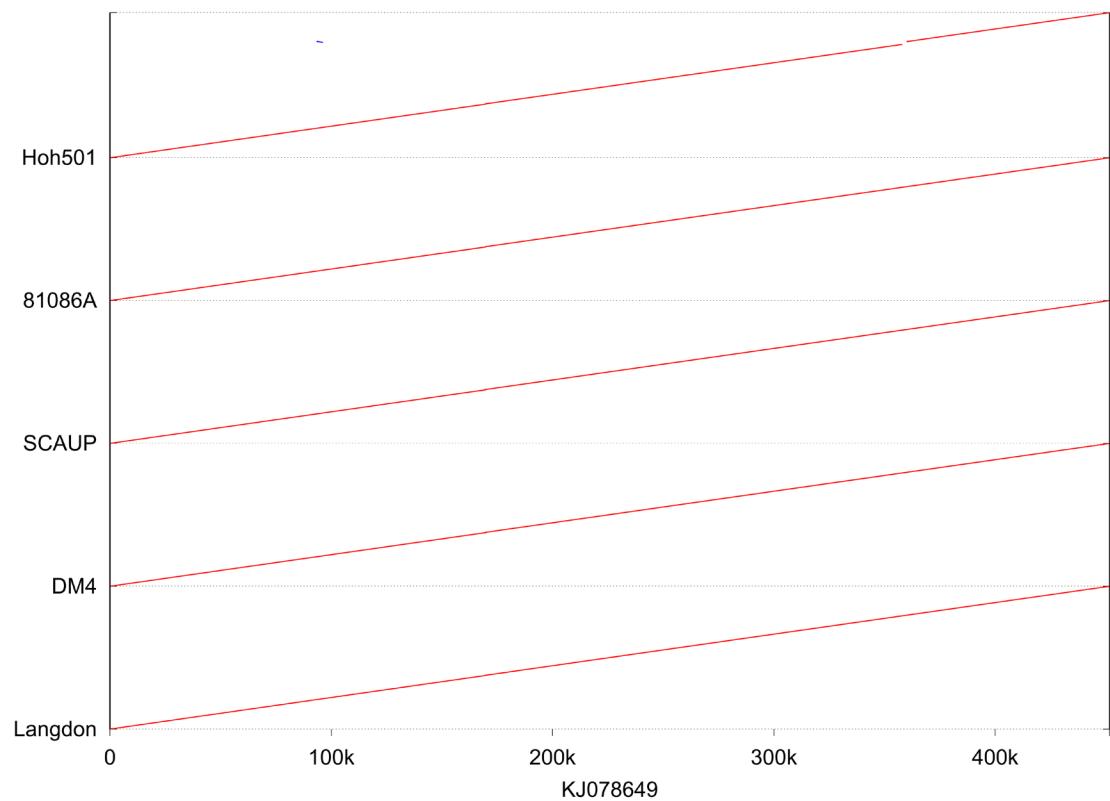


Fig. S11. **MUMmerplot of assembled mitochondrial genomes in tetraploid wheats.** *Triticum dicoccum* SCAUP, *T. dicoccum* DM4, *T. turgidum* Langdon, *T. turgidum* 81086A, *T. turgidum* Hoh501; KJ078649: *T. durum* cell-line 56-1.

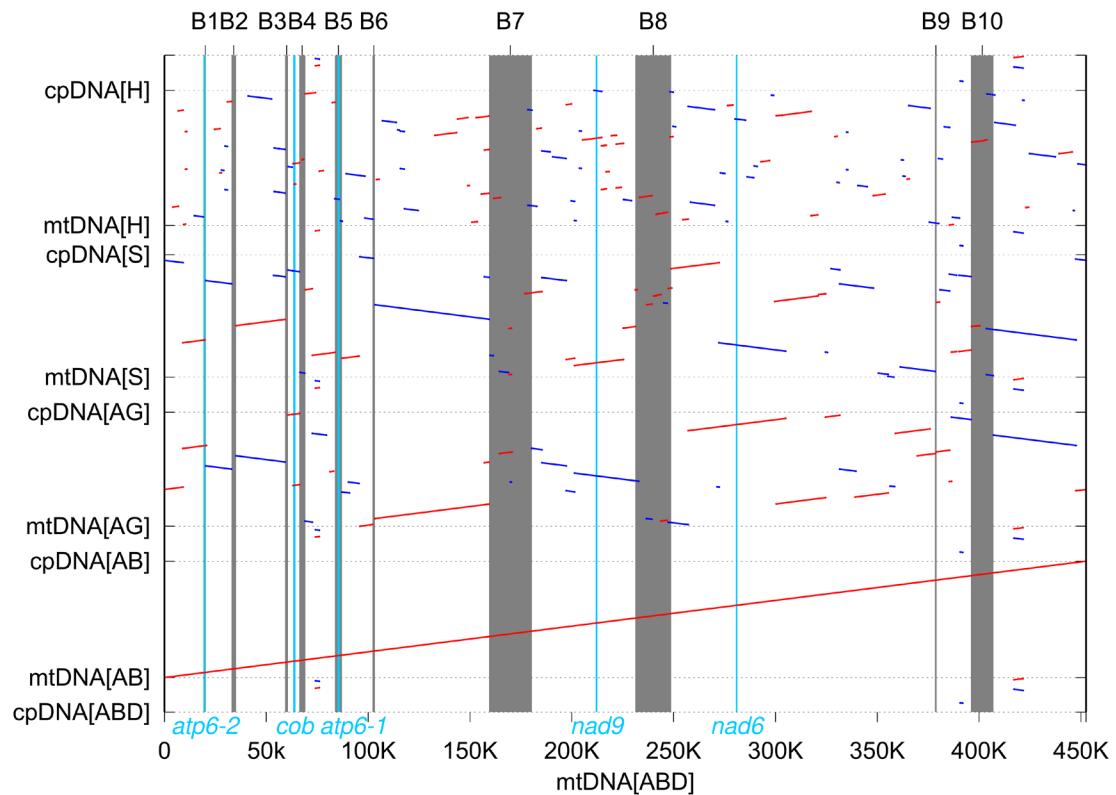


Fig. S12. Sequence synteny to *Triticum aestivum* mitochondrial genome. cpDNA: chloroplast genome, mtDNA: mitochondrial genome; ABD: *Triticum aestivum*; AB: *T. turgidum*; AG: *T. timopheevii*; S: *Aegilops speltoides*; H: *Hordeum vulgare*; Accession number: cpDNA [ABD] NC_002762, cpDNA [AB] NC_024814, mtDNA[AG] AP013106, cpDNA[AG] NC_024764, mtDNA[S] AP013107, cpDNA[S] NC_022135, mtDNA[H] MN127982, cpDNA[H] NC_008590. A total of 10 blocks (B1-B10, marked by grey color) of DNA were supposed to be transferred from BB nuclear genome, supported by BLAST hits. Among of them, 4 blocks brought new genes to mtDNA, such as *atp6-2* in B1, *atp6-1* in B5, *orf359* in B6, *trnK+trnQ+rnn26-p* in B7, and *atp8-2* in B8; And B1 and B5 seems independently occurred and made the *atp6* doubled and unique in the *Triticum-Aegilops* complex.

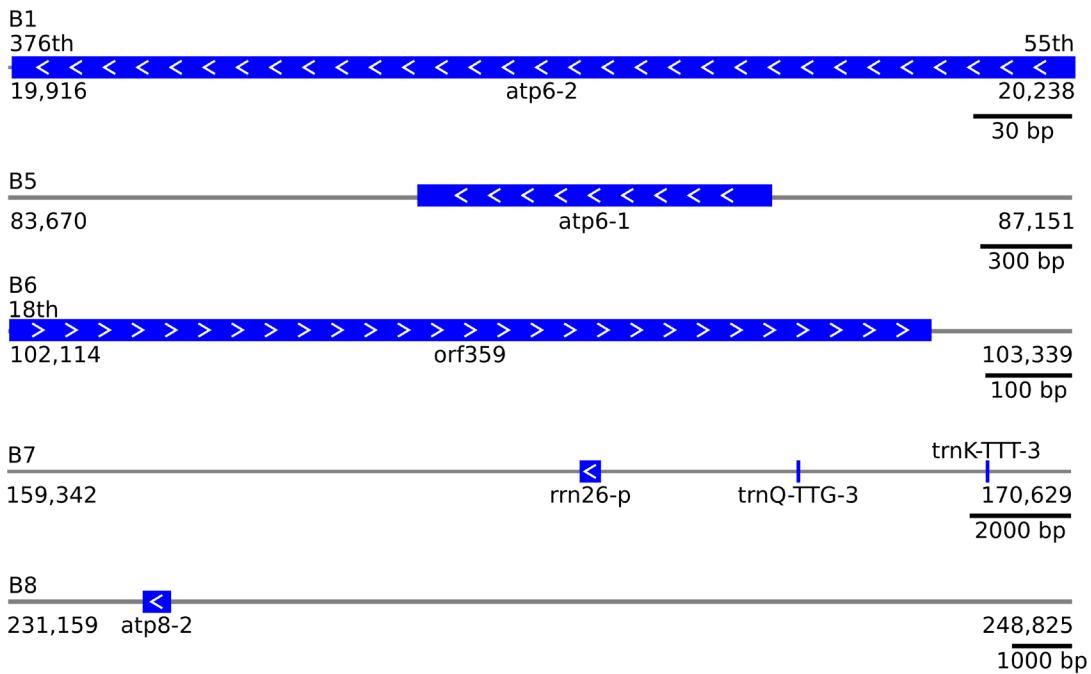


Fig. S13. Mitochondrial genes supposed to be from the BB nuclear genome. The original *atp6* gene in Block1 was destroyed due to an mtDNA re-arrangement, and then repaired by a nuclear version; And the Block5 brought another identical copy of nuclear *atp6*, although we have no solid evidence to figure out which one occurred first. The *atp8-2* in Block8 was have 5 unique SNPs, which could be differentiated from the original mtDNA *atp8-1* gene.

Schematic of Mitochondrial ATP6 Proteins

ATP synthase A = Pfam PF00119
 Motif = MEME Motif

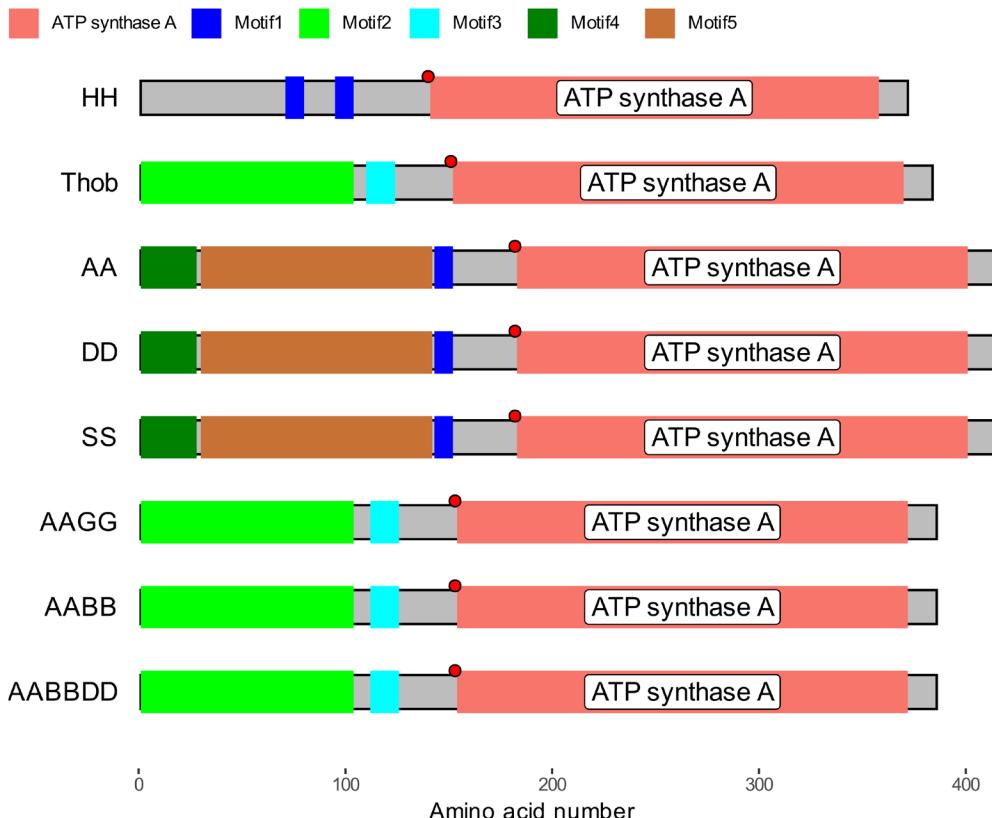


Fig. S14. Domain and motif analysis of the mitochondrial ATP6 proteins. The motifs or conserved domains were highlighted by different colors. The red circles marked the border of conserved C-terminal sequences, while the N-terminals were rather varied in ATP6, which were predicted to be outside the membrane for a membrane-bound protein. HH: *Hordeum vulgare*; Thob: *Thinopyrum obtusiflorum*; AA: *Triticum urartu*; DD *Aegilops tauschii*; SS: *Ae. speltoides*; AAGG: *T. aestivum*; AABB: *T. turgidum*; AABBDD: *T. aestivum*.

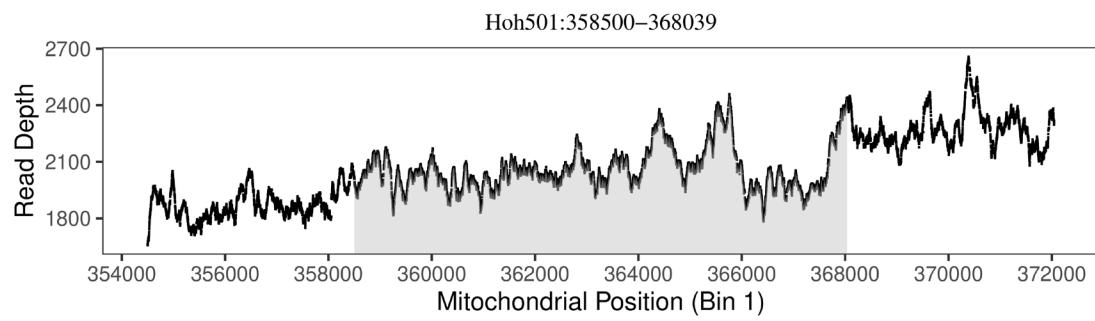


Fig. S15. Read depth around the nuclear DNA insertion in *Triticum turgidum* Hoh501's mtDNA.
The read depth of the insertion region, indicated by grey color, was consistent with that of the flanking regions, supporting that it was not a contamination from nuclear genome; Base resolution was set to 1 (Bin size).

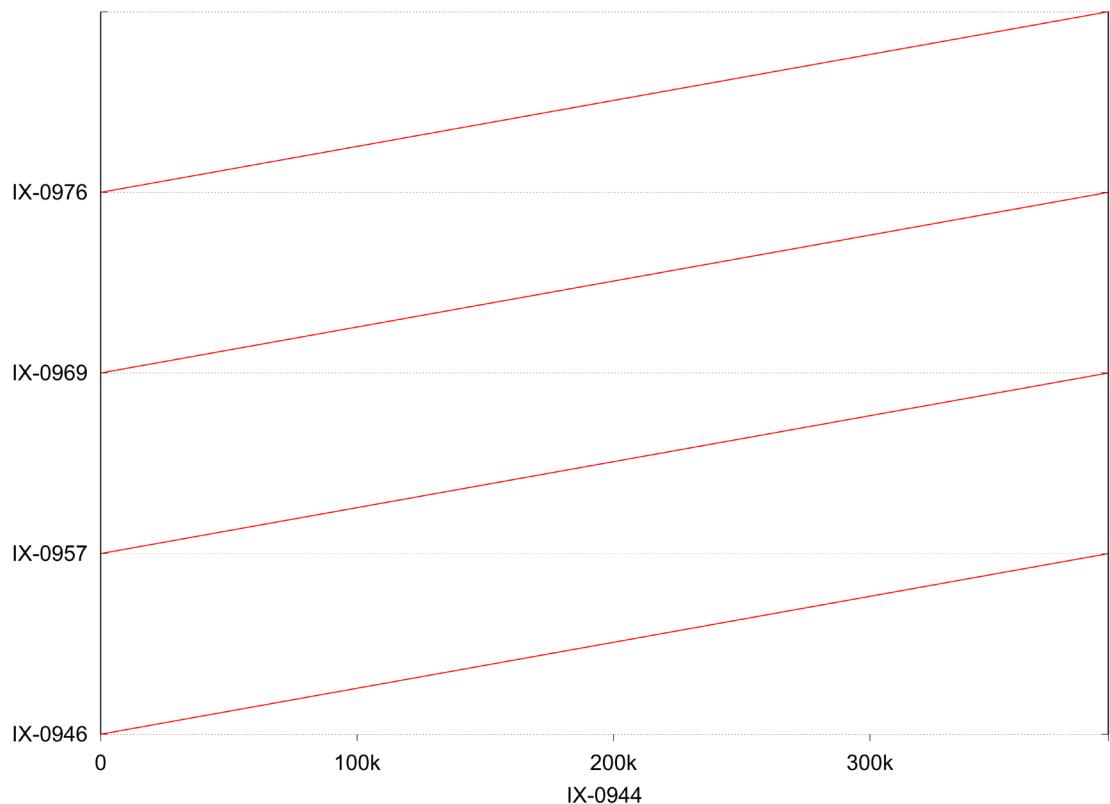


Fig. S16. MUMmerplot of assembled mitochondrial genomes in *Triticum urartu*. Five mtDNAs of *T. urartu* (AA) lines were assembled in this study. The mtDNAs and genic regions were quite conserved in AA.

```

          ..... 10 ..... 20 ..... 30 ..... 40 ..... 50 ..... 60
CCDD      ATGACTATAAGGAACCAACGATTCTCTCTTAAACAACCTATATACTCCACACTTAAC
SS        ATGACTATAAGGAACCAACGATTCTCTCTTAAACAACCAATATAACGGAACACTTAAC
AAGG      ATGACTATAAGGAACCAACGATTCTCTCTTAAACAACCTATATACTCCACACTTAAC
AABB      ATGACTATAAGGAACCAACGATTCTCTCTTAAACAACCTATATACTCCACACTTAAC
AABBDD    ATGACTATAAGGAACCAACGATTCTCTCTTAAACAACCTATATACTCCACACTTAAC
consensus ATGACTATAAGGAACCAACGATTCTCTCTTAAACAACCTATATAACtccACACTTAAC
          ..... 1150 ..... 1160 ..... 1170 ..... 1180 ..... 1190 ..... 1200
CCDD      GGAATTCCAAAATATTACACGGATGAGACTCATCGCACCGGATCCTTTCTCCTTTGGC
SS        GGAATTCCAAAATATTACACGGATGAGACTCATCGCACCGGATCCTTTCTCCTTTGGC
AAGG      GGAATTCCAAAATATTACACGGATGAGACTCATCGCACCGGATCCTTTCTCCTTTGGC
AABB      GGAATTCCAAAATATTACACGGATGAGACTCATCGCACCGGATCCTTTCTCCTTTGGC
AABBDD    GGAATTCCAAAATATTACACGGATGAGACTCATCGCACCGGATCCTTTCTCCTTTGGC
consensus GGAATTCCAAAATATTACACGGATGAGACTCATCGCACCGGATCCTtttcctTTTGGC
          ..... 1210 ..... 1220 ..... 1230 ..... 1240 ..... 1250 ..... 1260
CCDD      CATAGCCATATAAGAACGTCCTCTCCGAAGCGGAGGGGAATCTCAGGATTCCCTAT
SS        CATAGCCATAGAACGACTGCTCTCCGAAGCGGAGGGAGGGGAATCTATTGATTCCCTAT
AAGG      CATAGCCATAGAACGACTGCTCTCCGAAGCGGAGGGAGGGGAATCTATTGATTCCCTAT
AABB      CATAGCCATAGAACGACTGCTCTCCGAAGCGGAGGGAGGGGAATCTATTGATTCCCTAT
AABBDD    CATAGCCATAGAACGACTGCTCTCCGAAGCGGAGGGAGGGGAATCTATTGATTCCCTAT
consensus CATAGCCATAGAACGACTGCTCTCCGAAGCGGAGGGAGGGGAATCTATTGATTCCCTAT
          ..... 1450 ..... 1460 ..... 1470 ..... 1480 ..... 1490 ..... 1500
CCDD      GAAGCGATG-----GAGGAGGAAACAGATACAA-----
SS        GAAGTAATGGCTAAAGAGGAGGAAACAGATACAAGCCTCCGGTGGAAAGGTTGTGAAGAC
AAGG      GAAGTAATGGCTAAAGAGGAGGAAACAGATACAAGCCTCCGGTGGAAAGGTTGTGAAGAC
AABB      GAAGTAATGGCTAAAGAGGAGGAAACAGATACAA-----
AABBDD    GAAGTAATGGCTAAAGAGGAGGAAACAGATACAA-----
consensus GAAGTAATGGCTAAAGAGGAGGAAACAGATACAA
          ..... 1510 ..... 1520 ..... 1530 ..... 1540 ..... 1550 ..... 1560
CCDD      CTAAGTATGACGGAGGAGGAAACAGATACAAAGCCTCCGGTGGAAAGGTTGTGAAAAGACC
SS        CTAAGTATGACGGAGGAGGAAACAGATACAAAGCCTCCGGTGGAAAGGTTGTGAAAAGACC
AAGG      CTAAGTATGACGG-----
AABB      -----
AABBDD    -----
consensus -----
          ..... 1570 ..... 1580 ..... 1590 ..... 1600 ..... 1610 ..... 1620
CCDD      -----GCCTCCGGTGGAAAGGTTGTGAAGACCGTC
SS        TAAGTATGACGGAGGAGGAAACAGATACAAAGCCTCCGGTGGAAAGGTTGTGAAGACCTAA
AAGG      -----AGGAGGAAACAGATACAAAGCCTCCGGTGGAAAGGTTGTGAAGACCTAA
AABB      -----GCCTCCGGTGGAAAGGTTGTGAAGACCTAA
AABBDD    -----GCCTCCGGTGGAAAGGTTGTGAAGACCTAA
consensus GCCTCCGGTGGAAAGGTTGTGAAGACCTaa
          ..... 1630 ..... 1640 ..... 1650 ..... 1660 ..... 1670 ..... 1680
CCDD      GTATGACTCCGGAGGAAAGGACAAAATTGCTGAAGCATTTCCTATGGGATAAAA
SS        GTATGACGGAGGAGGAAAGGAAAAAAATTGCTGGAAGCATTTCCTATGGGATAAAA
AAGG      GTATGACGGAGGAGGAAAGGAAAAAAATTGCTGGAAGCATTTCCTATGGGATAAAA
AABB      GTATGACGGAGGAGGAAAGGAAAAAAATTGCTGGAAGCATTTCCTATGGGATAAAA
AABBDD    GTATGACGGAGGAGGAAAGGAAAAAAATTGCTGGAAGCATTTCCTATGGGATAAAA
consensus GTATGACGGAGGAGGAAAGGAAaaaATTGCTGAAGCATTTCCTATGGGATAAAA
          .....
CCDD      AAAAAATGA
SS        AACAAATGA
AAGG      AACAAATGA
AABB      AACAAATGA
AABBDD   AACAAATGA
consensus AACAAATGA

```

Fig. S17. Long versions of mitochondrial *cob* genes in *Triticum/Aegilops* complex. Some conserved alignment blocks were not shown here. The start codons were marked as green and stop codons as red color. Compared to the mtDNA *cob* genes in *Triticum urartu* (AA), *Aegilops tauschii* (DD) and *Hordeum vulgare*, these *cob* genes got additional 3' end sequences (alignment position 1,188-1,628) and were supposed to transfer from nuclear genomes. The sequence variations of SNPs and insertions-deletions may suggest independent stories of gene exchange between the mtDNA and the nuclear genome in *Ae. cylindrica* (CCDD), *Ae. speltoides*, *T. timopheevii* (AAGG) and the BB ancestor of *T. turgidum* (AABB) and *T. aestivum* (AABBDD)

Schematic of Mitochondrial COB Proteins

Cytochrome_b = Cytochrome_b, PF00033
 cob_C = Cytochrome_b (C-terminal), PF00032
 Motif = MEME Motif

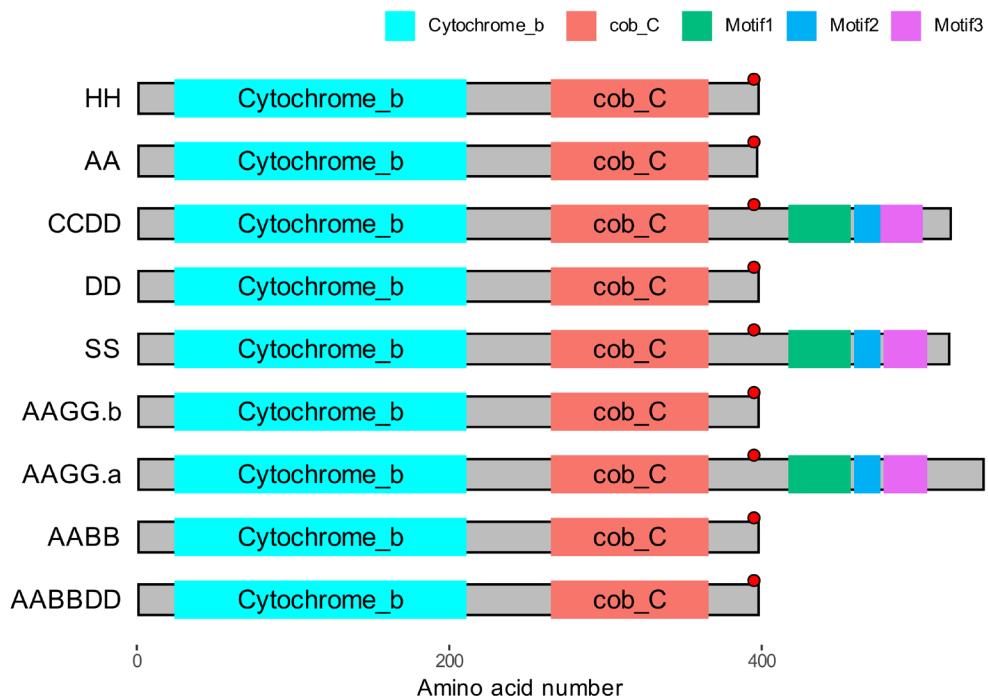


Fig. S18. Domain and motif analysis of the mitochondrial Cytochrome-B proteins. The motifs or conserved domains were highlighted by different colors. The red circles marked the border of conserved N-terminal sequences, while several rearrangements made the C-terminal diverse in *Aegilops cylindrica* (CCDD), *Ae. speltoides* (SS), *Triticum aestivum* (AAGG); The *cob* genes in *T. turgidum* (AABB) and *T. aestivum* (AABBDD) were actually a longer version, but retained a short ORF due to a nonsense mutation. Two *cob* genes were found in AAGG's mtDNA: long version *cob-A* and a short version *cob-B*. HH: *Hordeum vulgare*; AA: *T. urartu*; DD: *Ae. tauschii*.

Schematic of Mitochondrial NAD6 Proteins

NADH_UbQ/plastoQ_OxRdtase = NADH:ubiquinone/plastoquinone oxidoreductase, chain 6, PF00499

█ NADH_UbQ/plastoQ_OxRdtase

Motif = MEME Motif █ Motif1 █ Motif2 █ Motif3 █ Motif4 █ Motif5

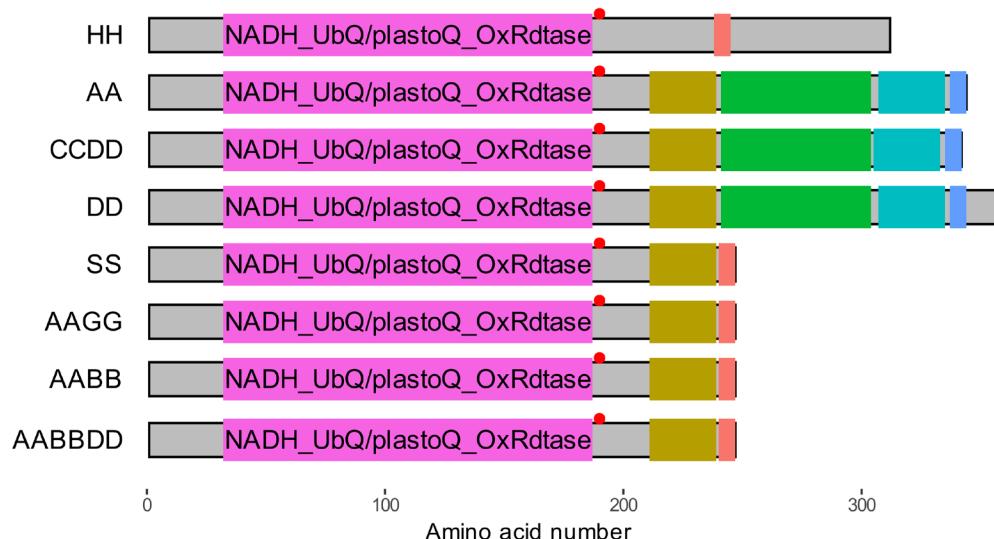


Fig. S19. Domain and motif analysis of the mitochondrial NAD6 proteins. The motifs or conserved domains were highlighted by different colors. The red circles marked the border of conserved N-terminal sequences, while several rearrangements made the C terminal diverse. The C-terminals were predicted to be a region outside the membrane for a membrane-bound protein. *Hordeum vulgare*; AA: *Triticum urartu*; DD *Aegilops tauschii*; SS: *Ae. speltoides*; AAGG: *T. aestivum*; AABB: *T. turgidum*; AABBDD: *T. aestivum*.

 10 20 30 40 50
HH . nad9 -1	ATGCTCTGTATAA T A C T T T C C C G A G C G A T G G T T A G C G G A T T C G G A A T
HH . nad9 -2	ATGCTCTGTATAA T A C T T T C C C G A G C G A T G G T T A G C G G A T T C G G A A T
AA . nad9	ATGCTCTGTATAA T A C T T T C C C G A G C G A T G G T T A G C G G A T T C G G A A T
CCDD . nad9	ATG CTCTGTATAA T A C T T T C C C G A G C G A T G G T T A G C G G A T T C G G A A T
DD . nad9	ATGCTCTGTATAA T A C T T T C C C G A G C G A T G G T T A G C G G A T T C G G A A T
SS . nad9	ATG CTCTGTATAA T A C T T T C C C G A G C G A T G G T T A G C G G A T T C G G A A T
AAGG . nad9	ATG CTCTGTATAA T A C T T T C C C G A G C G A T G G T T A G C G G A T T C G G A A T
AABB . nad9	ATG CTCTGTATAA T A C T T T C C C G A G C G A T G G T T A G C G G A T T C G G A A T
AABBDD . nad9	ATG CTCTGTATAA T A C T T T C C C G A G C G A T G G T T A G C G G A T T C G G A A T
consensus	ATGCTCTGTATAA T A C T T T C C C G A G C G A T G G T T A G C G G A T T C G G A A T
 110 120 130 140 150
HH . nad9 -1	GCCGATCCTGGATACATAACTCTAAAAGTG-----TGCAGTGAGGGATCT
HH . nad9 -2	GCCGATCCTGGATACATAACTCTAAAAGTG-----TGCAGTGAGGGATCT
AA . nad9	GCCGATCCTGGATACATAACTCTAAAAGTG-----TGCAGTTGGATCT
CCDD . nad9	GCCGATCCTGGATACAT C ACTCTAAAAGTG-----TGCAGTTGGATCT
DD . nad9	GCCGATCCTGGATACATAACTCTAAAAGTG-----TGCAGTTGGATCT
SS . nad9	GCCGATCCTGGATACATAACTCTAAAAGTG-----TGCAGTTGGATCT
AAGG . nad9	GCCGATCCTGGATACATAACTCTAAAAGTG-----TGCAGTTGGATCT
AABB . nad9	GCCGATCCTGGATACATAACTCTAAAAGTG-----TGCAGTTGGATCT
AABBDD . nad9	GCCGATCCTGGATACATAACTCTAAAAGTG-----TGCAGTTGGATCT
consensus	GCCGATCCTGGATACATAACTCTaaaAAAGTgt aTGAGTtttGGATCT
 260 270 280 290 300
HH . nad9 -1	CAAGAACCGGAACCAAAATAAAGCTTCTTATTTCATTT ATGG GATAAC
HH . nad9 -2	CAAGAACCGGAACCAAAATAAAGCTTCTTATTTCATTT ATGG GATAAC
AA . nad9	CAAGAACCGGAACCAAAATAAAGCTTCTTATTTCATTT ATGG GATAAC
CCDD . nad9	CAAGAACCGGAACCAAAATAAAGCTTCTTATTTCATTT ATGG GATAAC
DD . nad9	CAAGAACCGGAACCAAAATAAAGCTTCTTATTTCATTT ATGG GATAAC
SS . nad9	CAAGAACCGGAACCAAAATAAAGCTTCTTATTTCATTT ATGG GATAAC
AAGG . nad9	CAAGAACCGGAACCAAAATAAAGCTTCTTATTTCATTT ATGG GATAAC
AABB . nad9	CAAGAACCGGAACCAAAATAAAGCTTCTTATTTCATTT ATGG GATAAC
AABBDD . nad9	CAAGAACCGGAACCAAAATAAAGCTTCTTATTTCATTT ATGG GATAAC
consensus	CAAGAACCGGAACCAAAATAAAGCTTCTTATTTCATTT ATGG GATAAC
 460 470 480 490 500
HH . nad9 -1	CAAGTTTCGATCGATATTGCGGAGTGGATCATCCCTCTCGAAAACG CAG
HH . nad9 -2	CAAGTTTCGATCGATATTGCGGAGTGGATCATCCCTCTCGAAAACG CAG
AA . nad9	CAAGTTTCGATCGATATTGCGGAGTGGATCATCCCTCTCGAAAACGAAG
CCDD . nad9	CAAGTTTCGATCGATATTGCGGAGTGGATCATCCCTCTCGAAAACGAAG
DD . nad9	CAAGTTTCGATCGATATTGCGGAGTGGATCATCCCTCTCGAAAAC CAG
SS . nad9	CAAGTTTCGATCGATATTGCGGAGTGGATCATCCCTCTCGAAAACGAAG
AAGG . nad9	CAAGTTTCGATCGATATTGCGGAGTGGATCATCCCTCTCGAAAACGAAG
AABB . nad9	CAAGTTTCGATCGATATTGCGGAGTGGATCATCCCTCTCGAAAACGAAG
AABBDD . nad9	CAAGTTTCGATCGATATTGCGGAGTGGATCATCCCTCTCGAAAACGAAG
consensus	CAAGTTTCGATCGATATTGCGGAGTGGATCATCCCTCTCGAAAAC GAG
 860
HH . nad9 -1	GTAGCGACGGATAA
HH . nad9 -2	GTAGCGACGGATAA
AA . nad9	GTAGCGACGGATAA
CCDD . nad9	GTAGCGACGGATAA
DD . nad9	GTAGCGACGGATAA
SS . nad9	GTAGCGACGGATAA
AAGG . nad9	GTAGCGACGGATAA
AABB . nad9	GTAGCGACGGATAA
AABBDD . nad9	GTAGCGACGGATAA
consensus	GTAGCGACGGATAA

Fig. S20. Sequence alignment of the mitochondrial *nad9* genes in *Triticum/Aegilops* complex. The 4-bp insertion (Alignment position: 132-135) altered the start codon and made *nad9* open reading frame longer at 5' end in *Aegilops speltoides* (SS), *Triticum timopheevii* (AAGG), *T. turgidum* (AABB) and *T. aestivum* (AABBDD). It was supposed to result from a homologous recombination with a nuclear counterpart. HH: *Hordeum vulgare*; AA: *T. urartu*; DD: *Ae. tauschii*.

Schematic of Mitochondrial NAD9 Proteins

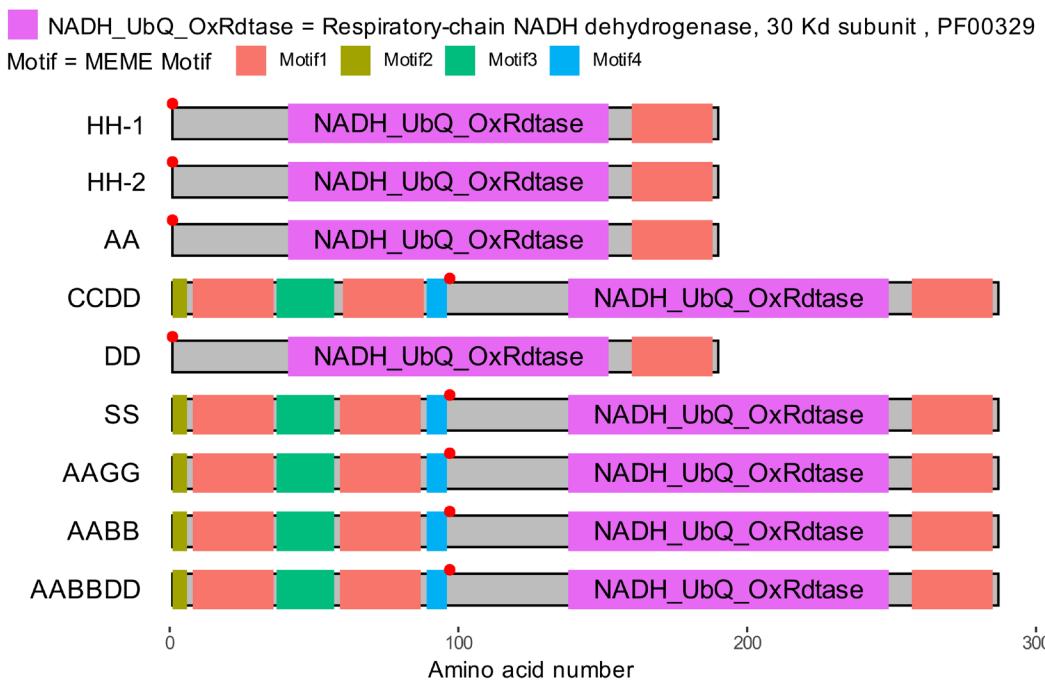


Fig. S21. Domain and motif analysis of the mitochondrial NAD9 proteins. The motifs or conserved domains were highlighted by different colors. The red circles marked the border of conserved C-terminal sequences, while the a 4-bp insertion in the upstream of *nad9* extended the open reading frame and brought new N-terminals in *Aegilops cylindrica* (CCDD), *Ae. speltoides* (SS), *Triticum timopheevii* (AAGG), *T. turgidum* (AABB) and *T. aestivum* (AABBDD). HH: *Hordeum vulgare*, two copies: *nad9-1* and *nad9-2*; AA: *T. urartu*; DD: *Ae. tauschii*.

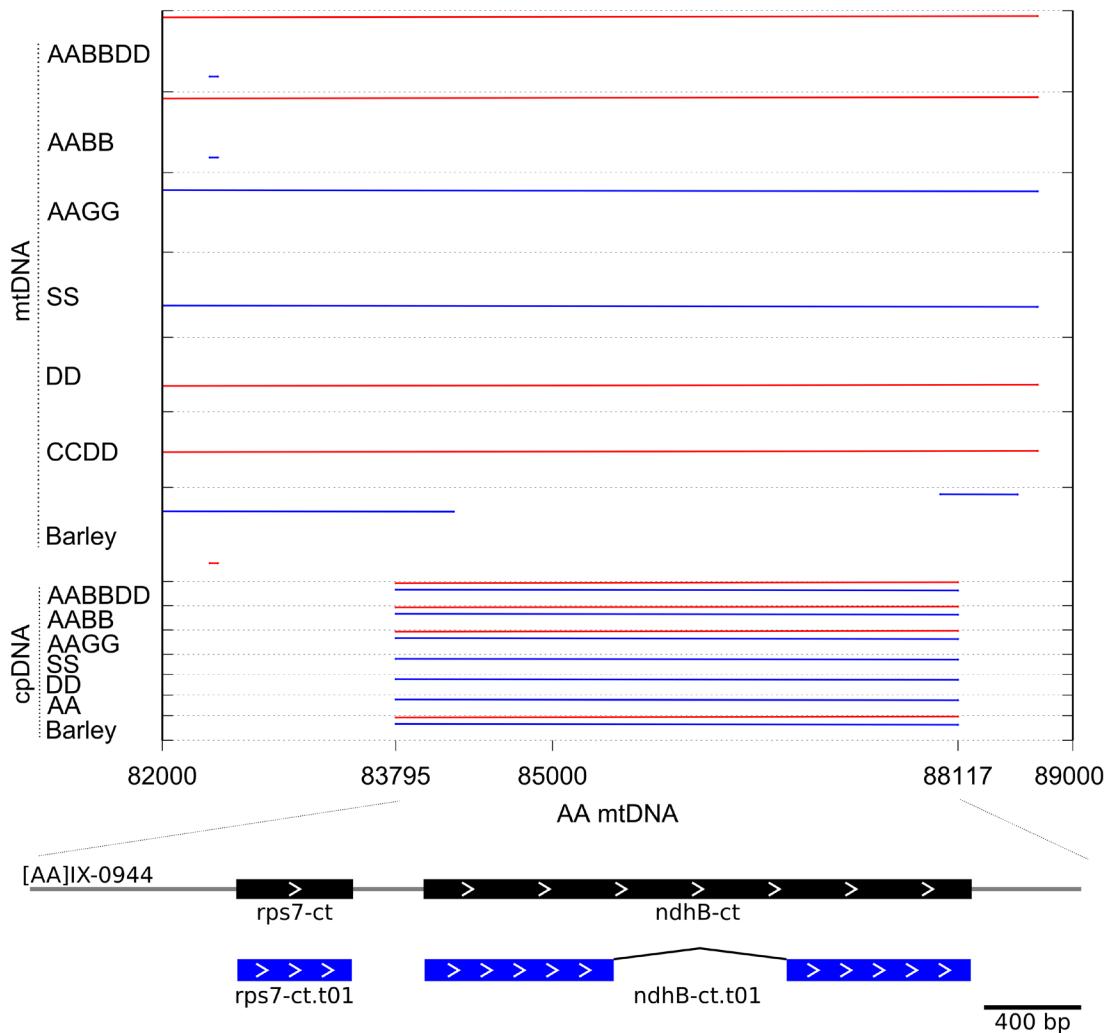


Fig. S22. Sequence alignment of *rps7-ct*–*ndhB-ct* fragments in Barley, and *Triticum/Aegilops*. This 4,323bp fragment has two copies in the barley's chloroplast genome (cpDNA), and one of them is supposed to be moved to mitochondrial genome (mtDNA), before the divergence to *Triticum urartu* (AA) or *Aegilops speltoides* (SS) species. So, all the subsequently diverged species retained this fragment in their mtDNAs. And this fragment in cpDNA got duplicated again in the BB and GG, indicating GG was actually closer to BB than SS. NCBI sequences: barley cpDNA NC_008590 and mtDNA MN127982, AA cpDNA NC_021762, *Ae. tauschii* (DD) cpDNA NC_022133, SS cpDNA NC_022135 and mtDNA AP013107, *T. timopheevii* (AAGG) cpDNA NC_024764 and mtDNA AP013106, *T. turgidum* AABB cpDNA NC_024814, *T. aestivum* (AABBDD) cpDNA NC_002762.

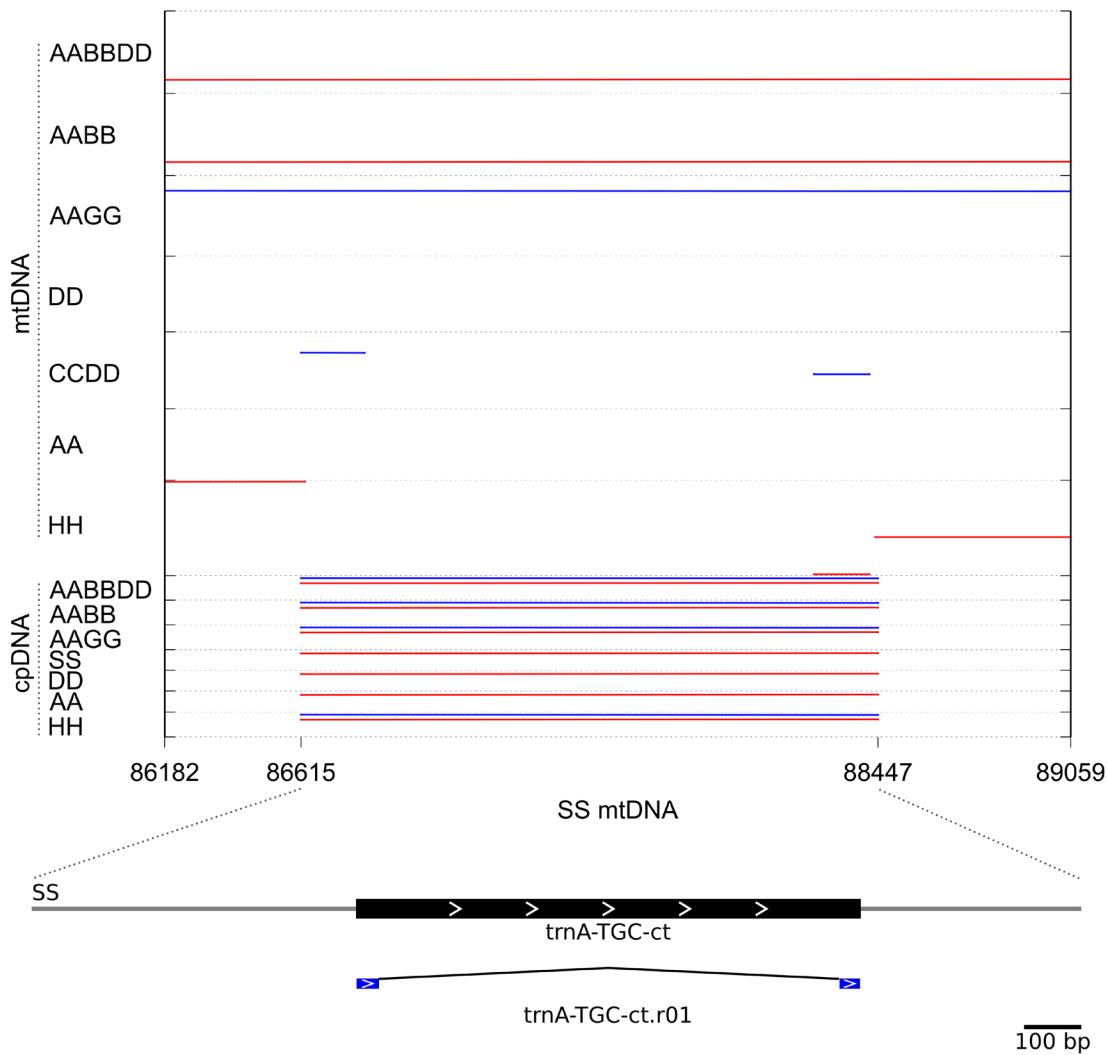


Fig. S23. Sequence alignment of the *trnA-ct* fragments in Barley and *Triticum-Aegilops*. This 1,833bp chloroplast (cpDNA) fragment, carrying *trnA-TGC-ct*, was copied to mitochondrial genomes (mtDNAs) of BB and GG. And due to a long-fragment duplication, this fragment got duplicated again in cpDNA of BB and GG, indicating GG is actually closer to BB than *Aegilops speltoides* (SS). NCBI sequences: barley cpDNA NC_008590 and mtDNA MN127982, AA cpDNA NC_021762, *Ae. tauschii* (DD) cpDNA NC_022133, SS cpDNA NC_022135 and mtDNA AP013107, *T. timopheevii* (AAGG) cpDNA NC_024764 and mtDNA AP013106, *T. turgidum* AABB cpDNA NC_024814, *T. aestivum* (AABBDD) cpDNA NC_002762.

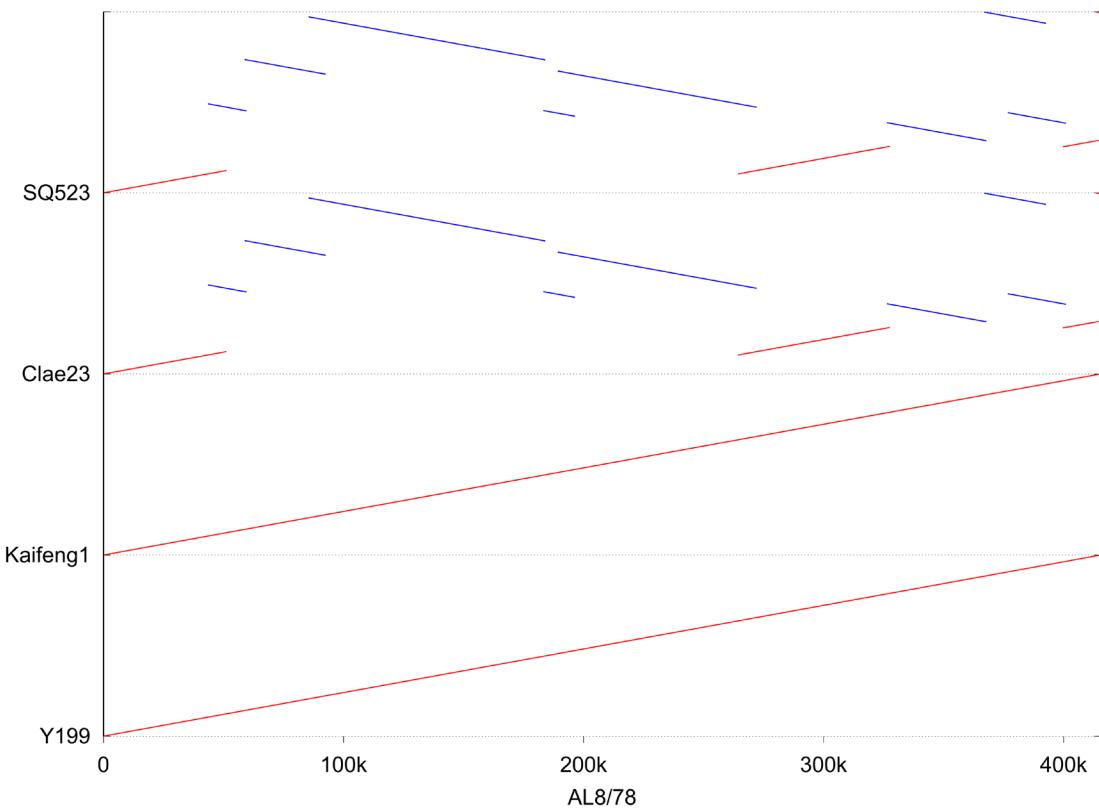


Fig. S24. **MUMmerplot of assembled mitochondrial genomes in *Aegilops tauschii*.** These 5 mtDNAs in the DD group were assembled in this study. At least 2 sequence re-arrangements in SQ523 and Clae23 were confirmed using distance between paired reads from Illumina sequencing,

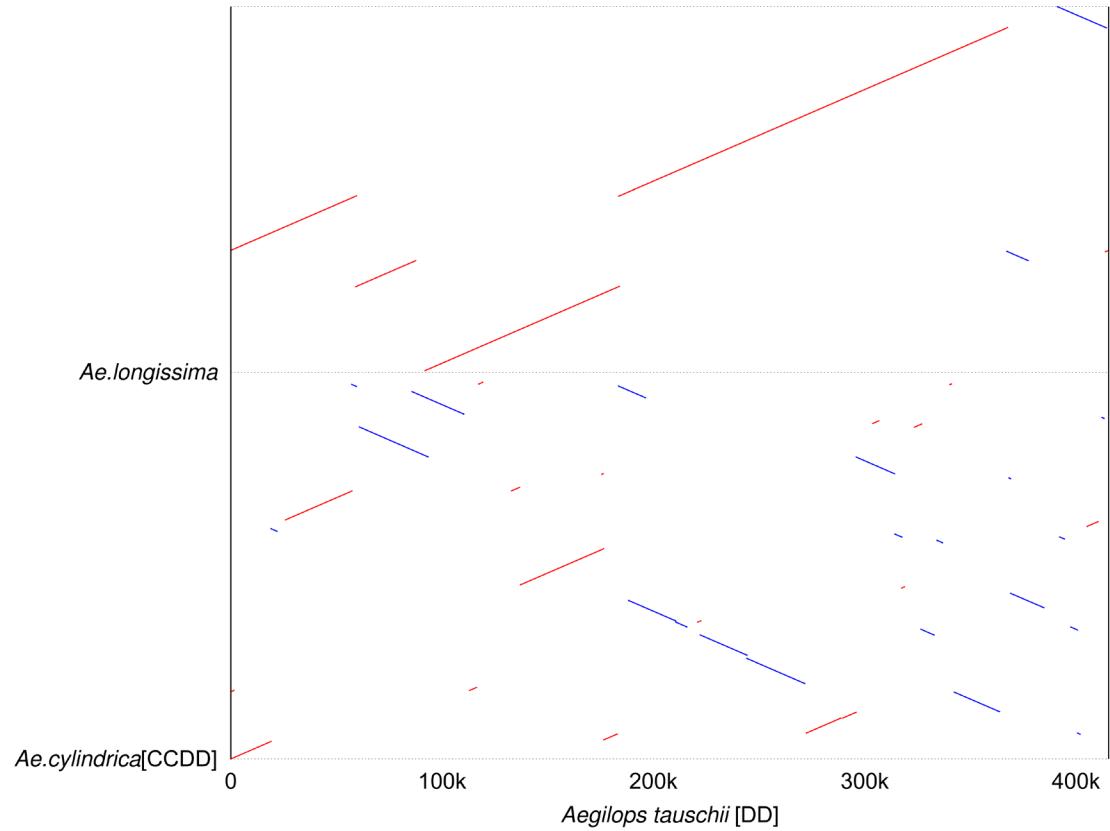


Fig. S25. mtDNA synteny of *Aegilops cylindrica* and *Ae. longissima* against *Ae. tauschii*. *Ae. cylindrica* (CCDD) Jinan20190529 and *Ae. tauschii* (DD) AL8/78's mtDNA genomes were assembled in this study, and *Ae. longissima* G759 (GenBank KJ078648) was downloaded from the NCBI Organelle Genome Resources database. The AL8/78 was of good synteny with *Ae. longissima* in overall despite of some potential sequence re-arrangements. The fragmented synteny of CCDD Jinan20190529 may reflect the differences between AL8/78 and its DD donor *Ae. tauschii* Coss.

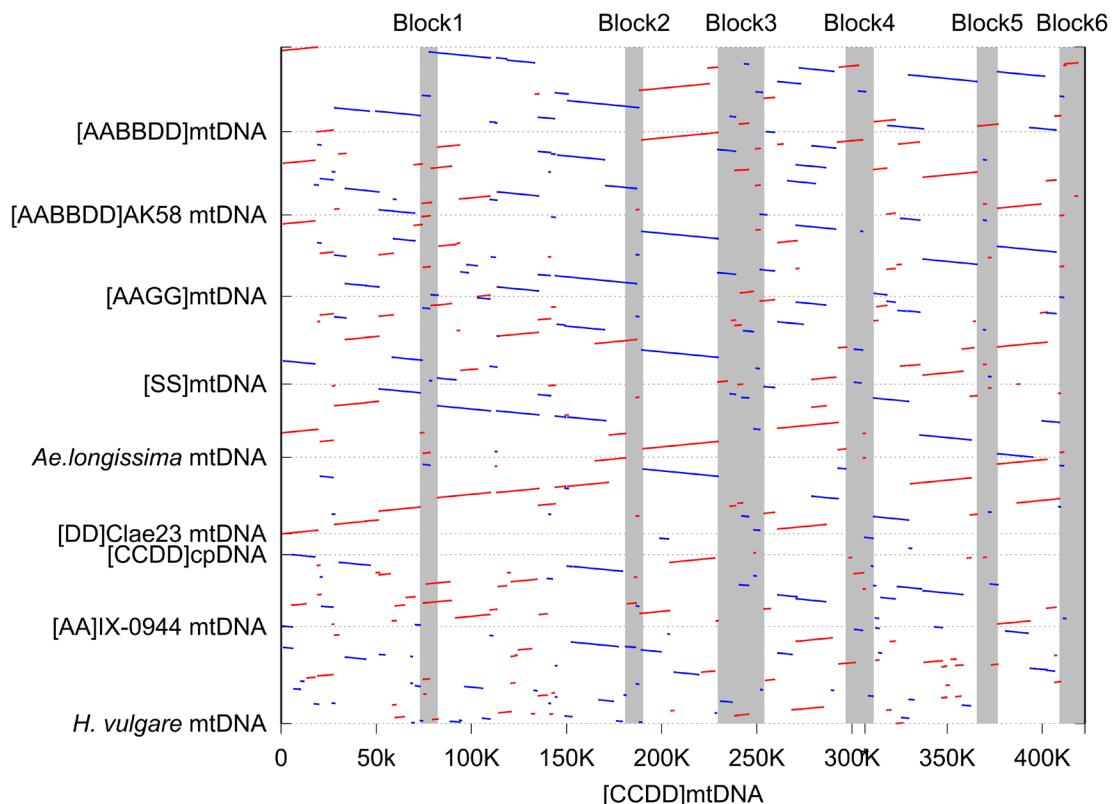


Fig. S26. Sequence synteny to *Aegilops cylindrica* mitochondrial genome. Block1-6 in *Aegilops cylindrica* (CCDD) mtDNA were supposed to be non-mtDNA sequences, which might be originated from chloroplast DNA (cpDNA), nuclear genome or horizontal gene transfer from other species like wheats. These foreign fragments brought new genes to CCDD, such as *cob* in Block1, *cox3* in Block2, *orf256* and *cox1* in Block3, *trnV-GAC* in Block4, *atp1-p* in Block5 and *orf359-p* in Block6. *Hordeum vulgare* mitochondrial DNA (mtDNA) MN127982; *Ae. cylindrica* chloroplast DNA (cpDNA) NC_023096; *Ae. longissima* mtDNA KJ078648; *Ae. speltoides* (SS) mtDNA AP013107; *Triticum timopheevii* (AAGG) mtDNA AP013106; *T. aestivum* Chinese Spring M°-type mtDNA AP013053.

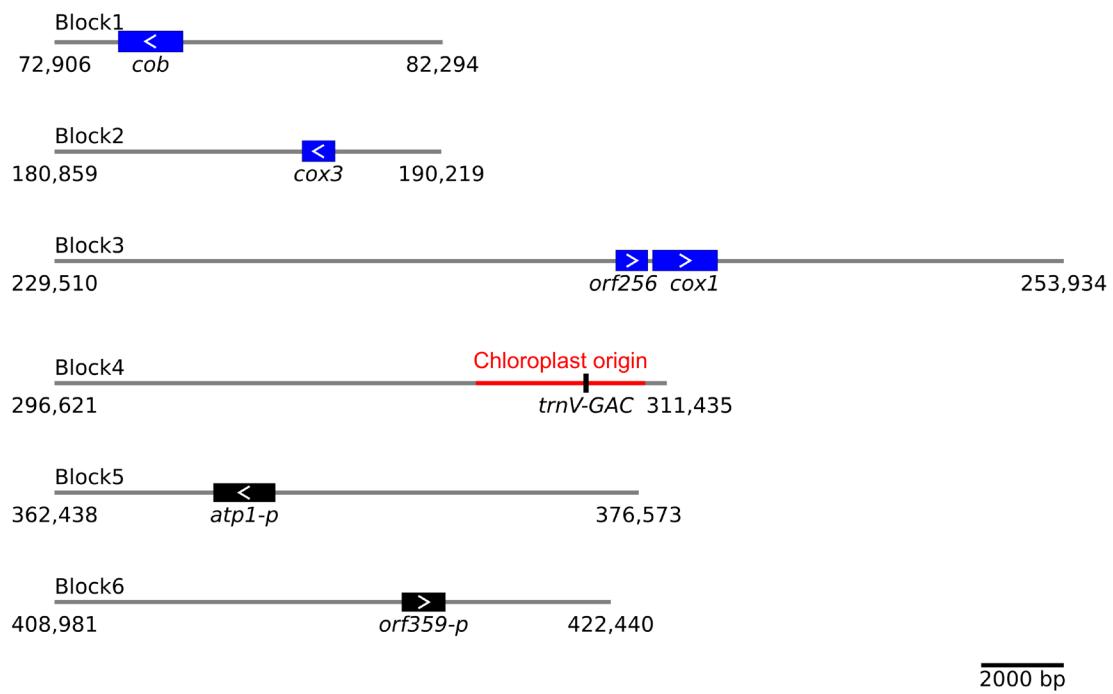


Fig. S27. Genes in six blocks of non-mtDNA regions in *Aegilops cylindrica*'s mtDNA. Most of these DNA sequences were supposed to be transferred from nuclear genome or other species like wheats. A small region (marked by red line) in Block4 was confirmed to be from the chloroplast genome of *Aegilops cylindrica* (CCDD). The *cob* in Block1 was longer than the *cob* in *Ae. tauschii* (DD) at 3' end, and was supposed to come from *Ae. markgrafii* (CC) nuclear genome. The *orf256-cox1* in Block3 was similar with those in *Ae. speloides* (SS) and *Triticum timopheevii* (AAGG). The *atp1-p* in Block5 was truncated at 3' end, and retained the same alleles with that in *T. urartu* (AA) and DD. The *orf359-p* in Block6 was truncated at 5' end and similar to that in SS and *Thinopyrum obtusiflorum*.

Table S1. NCBI sequences for mtDNAs and cpDNAs in wheat-related species

Species names	Symbols	mtDNAs	cpDNAs
<i>Hordeum vulgare</i>	HH	MN127982	NC_008590
<i>Thinopyrum obtusiflorum</i>	-	OK120846	-
<i>Elymus sibiricus</i>	-	MZ202552	NC_058919
<i>Triticum urartu</i>	AA	-	NC_021762
<i>Aegilops longissima</i>	-	KJ078648*	NC_024830
<i>Aegilops cylindrica</i>	CCDD	-	NC_023096
<i>Aegilops tauschii</i>	DD	-	NC_022133
<i>Aegilops speltoides</i>	SS	NC_022666 / AP013107	NC_022135
<i>Triticum timopheevii</i>	AAGG	NC_022714 / AP013106	NC_024764
<i>Triticum turgidum</i>	AABB	KJ078649*	NC_024814
<i>Triticum aestivum</i>	AABBDD	NC_007579 / AP008982 NC_036024 / EU534409	NC_002762

* Some genic regions got assembly errors

Table S2. BLAST summary of the mitochondrial *atp6* gene of *Triticum aestivum*

Species	Chr	Identity (%)	Length (bp)	Start	End
<i>Hordeum vulgare</i> (HH)	3H	95.80	785	374,278,966	374,279,749
<i>Triticum urartu</i>	5	98.98	784	108,702,198	108,701,415
<i>Aegilops tauschii</i>	Chr5D	92.72	1,030	401,555,286	401,556,311
<i>Aegilops speltoides</i> *	Chr4S	96.47	1,161	261,457,037	261,455,878
<i>Triticum turgidum</i> **	2A	99.66	1,161	193,278,145	193,276,985
	7B	99.48	1,161	250,760,307	250,759,147
	2B	99.48	1,161	608,911,002	608,909,844
<i>Triticum aestivum</i> **	Chr7B	100.00	1,161	254,292,234	254,291,074
	Chr6B	99.74	1,161	134,726,611	134,725,451
	Chr5B	99.66	1,161	714,769,689	714,768,529

* Complete open reading frame (ORF) was not found in the Chr4S region of *Aegilops speltoides* (SS);

** Five out of 18 BLAST regions got complete ORFs in AABB/AABBDD.

Table S3. BLAST summary of the mitochondrial *cob* gene of *Aegilops cylindrica*

Species	SeqID	Identity %	Qstart*	Qend*	Sstart*	Send*
<i>Hordeum vulgare</i>	chr3H	99.75	1	1,187	224,030,100	224,028,915
<i>Triticum urartu</i>	chr1A	99.59	1	1,216	8,332,828	8,331,612
<i>Aegilops tauschii</i>	chr3D	98.15	1	1,187	510,188,976	510,190,160
<i>Aegilops speltoides</i>	Chr2S	94.30	1	1,516	413,090,985	413,092,482
<i>Triticum turgidum</i>	chr2A	96.95	1	1,566	418,056,173	418,054,604
	chr2B	96.88	1	1,566	784,456,032	784,454,470
<i>Triticum aestivum</i>	Chr2A	96.82	1	1,566	427,545,701	427,544,131
	Chr3A	96.76	1	1,566	239,428,419	239,429,986
	Chr5B	95.93	1	1,566	34,800,206	34,801,775

* Qstart=Query start; Qend=Query end; Sstart=Subject start; Send=Send

Table S4. BLAST summary of the mitochondrial *nad6* gene of *Triticum aestivum*

Species	Chr	Length (bp)	Identity (%)	Start	End
<i>Hordeum vulgare</i>	5H	745	93.96	292,313,342	292,314,075
<i>Triticum urartu</i>	7	742	97.17	563,666,562	563,667,301
<i>Aegilops tauschii</i>	Chr2D	742	96.36	290,020,894	290,021,634
<i>Aegilops speltoides</i>	Chr3S	690	96.52	283,141,557	283,142,246
<i>Triticum turgidum</i>	5A	744	100.00	624,156,010	624,156,753
<i>Triticum aestivum</i>	Chr2B	744	100.00	399,838,547	399,839,290
	Chr3B	744	100.00	185,638,576	185,639,319
	Chr3B	744	100.00	49,428,320	49,429,063
	Chr6B	744	100.00	725,368,467	725,369,210

Table S5. BLAST summary of 10 blocks of mtDNA sequences in *Triticum aestivum*

Block No.	Length (bp)	Genomes				
		<i>Triticum aestivum</i>	<i>Triticum turgidum</i>	<i>Aegilops speltoides</i>	<i>Triticum urartu</i>	<i>Hordeum vulgare</i>
Block1	323	1B	1B	4S	-	-
Block2	2,319	6B	1A	2S	-	-
Block3	1,491	2A	6B	1S	-	-
Block4	2,996	5B	5B	-	2	-
Block5	3,482	5B	2A	-	-	-
Block6	1,226	1B	1B	1S	5	-
Block7	25,995	Un*	1B	-	-	-
Block8	17,666	7A/4B	4B	2S	-	-
Block9	802	2A/2B/5B	2A/2B/3A	7S/3S	-	1H/2H
Block10	11,015	6B	6B	7S	-	-

* Un: assembled sequences not anchored to known chromosomes

References

- [1] M. Kokot, M. Długosz, S. Deorowicz, KMC 3: counting and manipulating k-mer statistics, *Bioinformatics* 33 (2017) 2759-2761. <https://doi.org/10.1093/bioinformatics/btx304>.
- [2] R. Chikhi, P. Medvedev, Informed and automated *k*-mer size selection for genome assembly, *Bioinformatics* 30 (2014) 31-37. <https://doi.org/10.1093/bioinformatics/btt310>.
- [3] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform, *Bioinformatics* 25 (2009) 1754 - 1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- [4] H. Thorvaldsdóttir, J.T. Robinson, J.P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, *Briefings in Bioinformatics* 14 (2013) 178-192. <https://doi.org/10.1093/bib/bbs017>.
- [5] Y. Ogihara, Y. Yamazaki, K. Murai, A. Kanno, T. Terachi, T. Shiina, N. Miyashita, S. Nasuda, C. Nakamura, N. Mori, et al, Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome, *Nucleic Acids Res.* 33 (2005) 6235-6250. <https://doi.org/10.1093/nar/gki925>.
- [6] P. Cui, H. Liu, Q. Lin, F. Ding, G. Zhuo, S. Hu, D. Liu, W. Yang, K. Zhan, A. Zhang, et al, A complete mitochondrial genome of wheat (*Triticum aestivum* cv. Chinese Yumai), and fast evolving mitochondrial genes in higher plants, *J. Genet.* 88 (2009) 299-307. <https://doi.org/10.1007/s12041-009-0043-9>.
- [7] IWGSC, Shifting the limits in wheat research and breeding using a fully annotated reference genome, *Science* 361 (2018) eaar7191. <https://doi.org/10.1126/science.aar7191>.
- [8] F. Ramírez, D.P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A.S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: a next generation web server for deep-sequencing data analysis, *Nucleic Acids Res.* 44 (2016) W160-W165. <https://doi.org/10.1093/nar/gkw257>.
- [9] T. Paysan-Lafosse, M. Blum, S. Chuguransky, T. Grego, B.L. Pinto, Gustavo A. Salazar, Maxwell L. Bileschi, P. Bork, A. Bridge, L. Colwell, et al, InterPro in 2022, *Nucleic Acids Res.* 51 (2022) D418-D427. <https://doi.org/10.1093/nar/gkac993>.
- [10] T.L. Bailey, M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li, W.S. Noble, MEME Suite: tools for motif discovery and searching, *Nucleic Acids Res.* 37 (2009) W202-W208. <https://doi.org/10.1093/nar/gkp335>.
- [11] P. Brennan, drawProteins: a Bioconductor/R package for reproducible and programmatic generation of protein schematics, *F1000Research* 7 (2018) 1105. <https://doi.org/10.12688/f1000research.14541.1>.
- [12] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. Madden, BLAST+: architecture and applications, *BMC Bioinf.* 10 (2009) 421. <https://doi.org/10.1186/1471-2105-10-421>.