RESEARCH ARTICLE

A complete mitochondrial genome of wheat (*Triticum aestivum* cv. Chinese Yumai), and fast evolving mitochondrial genes in higher plants

PENG CUI[1,5†], HUITAO LIU[2,5†], QIANG LIN[1,5], FENG DING[1,5], GUOYIN ZHUO[2,4], SONGNIAN HU[1], DONGCHENG LIU[2], WENLONG YANG[2], KEHUI ZHAN[3], AIMIN ZHANG[2*] and JUN YU[1*]

[1]*The CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, People's Republic of China*
[2]*The State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, People's Republic of China*
[3]*Agronomy College of Henan Agricultural University, Zhengzhou 450002, People's Republic of China*
[4]*College of Biological Sciences of China Agricultural University, Beijing 100094, People's Republic of China*
[5]*Graduate School of Chinese Academy of Sciences, Beijing 100049, People's Republic of China*

Abstract

Plant mitochondrial genomes, encoding necessary proteins involved in the system of energy production, play an important role in the development and reproduction of the plant. They occupy a specific evolutionary pattern relative to their nuclear counterparts. Here, we determined the winter wheat (*Triticum aestivum* cv. Chinese Yumai) mitochondrial genome in a length of 452 and 526 bp by shotgun sequencing its BAC library. It contains 202 genes, including 35 known protein-coding genes, three rRNA and 17 tRNA genes, as well as 149 open reading frames (ORFs; greater than 300 bp in length). The sequence is almost identical to the previously reported sequence of the spring wheat (*T. aestivum* cv. Chinese Spring); we only identified seven SNPs (three transitions and four transversions) and 10 indels (insertions and deletions) between the two independently acquired sequences, and all variations were found in non-coding regions. This result confirmed the accuracy of the previously reported mitochondrial sequence of the Chinese Spring wheat. The nucleotide frequency and codon usage of wheat are common among the lineage of higher plant with a high AT-content of 58%. Molecular evolutionary analysis demonstrated that plant mitochondrial genomes evolved at different rates, which may correlate with substantial variations in metabolic rate and generation time among plant lineages. In addition, through the estimation of the ratio of non-synonymous to synonymous substitution rates between orthologous mitochondrion-encoded genes of higher plants, we found an accelerated evolutionary rate that seems to be the result of relaxed selection.

[Cui P., Liu H., Lin Q., Ding F., Zhuo G., Hu S., Liu D., Yang W., Zhan K., Zhang A. and Yu J. 2009 A complete mitochondrial genome of wheat (*Triticum aestivum* cv. Chinese Yumai), and fast evolving mitochondrial genes in higher plants. *J. Genet*. **88**, 299–307]

Introduction

The mitochondria as an endosymbiont for almost all eukaryotes, have undergone reductive genome evolution because of a long evolutionary history of intracellular lifestyle. The endocellular endosymbionts obtain metabolic precursors synthesized by their hosts and have lost many of their own genes for biosynthetic pathways relative to their free-living cousins (Gray *et al*. 1999; Itoh *et al*. 2002). Plant mitochondrial genomes have evolved to meet specific demands of physiological and biochemical functions of photosynthetic organisms, through different strategies for genome structures, gene content, and DNA mutation rate from those of animals and unicellular eukaryotes. The length of plant mitochondrial genomes varies widely, ranging from 105 to 704 kb, largely due to the expansion and reduction of intergenic re-

*For correspondence. E-mail: junyu@big.ac.cn; amzhang@genetics.ac.cn.
†These authors contributed equally for this work.

Keywords. wheat; mitochondrial genome; evolution; relaxed selection.

gions, such as the insertion or deletion of non-mitochondrial sequences from the nuclear and chloroplast genomes, as well as viral DNAs. In addition, mitochondrial genomes have variable sub-genomic structures in higher plants, resulting in frequent intramolecular recombination mediated by their abundant repetitive sequences. Plant mitochondrial genomes generally encode 22 subunits of oxidative phosphorylation mechanism, six to eleven ribosomal proteins, 17 to 22 tRNA and three rRNA genes as well as numerous functionally-undefined open reading frames (ORFs) (Kubo *et al*. 1993; Itoh *et al*. 2002; Notsu *et al*. 2002; Handa 2003; Ogihara *et al*. 2005). A remarkable feature is that the plant mitochondrion-encoded genes (MEGs) have a much lower evolutionary rate as compared to that of the animals—the mutation rate estimated in plants is ~1/20 of the nucleus-encoded genes (NEGs) in the same lineages. In a sharp contrast, the rate among animals is ~10 times that of the NEGs aside from the drastic size differences between animal (~17 kb in length) and plant mitochondrial genomes (Lynch *et al*. 2006).

Plant mitochondrial genomes, encoding necessary proteins involved in the system of energy production, play an important role in the development and reproduction of the plant. They occupy a specific evolutionary pattern relative to nuclear counterparts (Lynch *et al*. 2006). To this date, there have been 10 complete mitochondrial genomes sequenced from higher plants, including one gymnosperm (cycas) and nine angiosperms (four dicotyledons: rapeseed, tabacum, sugar beet and *Arabidopsis*; five monocotyledons: wheat, rice, maize, sorghum, and *Tripsacum*). In addition, two complete mitochondrial genomes were determined for liverwort and moss (Oda *et al*. 1992; Kubo *et al*. 2000; Notsu *et al*. 2002; Handa 2003; Clifton *et al*. 2004; Ogihara *et al*. 2005; Sugiyama *et al*. 2005; Terasawa *et al*. 2007; Chaw *et al*. 2008). These sequence data provide an opportunity for the understanding of evolutionary signatures among mitochondrial genomes of higher plants.

In this study, we sequenced the complete mitochondrial genome of the winter wheat (*Triticum aestivum* cv. Chinese Yumai), and described the genome features with respect to the gene content and organization, codon usage and nucleotide composition. Moreover, through comparative analysis with other higher plant mitochondrial genomes, we investigated the evolutionary patterns of mitochondrial DNA (mtDNA) genes among higher plants.

Materials and methods

Specimens and mitochondrial DNA extraction

Triticum aestivum cv. Chinese Yumai is a winter wheat planted in He Nan, People's Republic of China. Mitochondria were isolated from etiolated two-weeks-old seedlings according to the procedure of Sugiyama *et al*. (2005). Mitochondrial fractions were collected by differential centrifugation, incubated with DNase I for 1 h on ice to eliminate free DNA released from broken organelles, and further purified by centrifugation in a discontinuous sucrose-density gradient (1.2 M/1.6 M/2.0 M). Purified mitochondrial bands were carefully collected from the 1.6 M/1.2 M interface and washed with 0.4 M sucrose. The fraction was finally lysed in 2% Sarkosyl for mtDNA extraction, followed by phenol-chloroform extraction and ethanol precipitation.

Genome library construction and sequencing

A mitochondrial genome BAC library was constructed following a previously published procedure (Osoegawa and de Jong 2004) with minor modifications. Mitochondrial genomic DNA was partly digested with *Sau3*AI, size-fractioned by pulsed-field gel electrophoresis, and ligated to PIndigoBAC-5 *Bam*HI cloning-ready vector (Epicentre Biotechnologies, Madison, USA; http://www.epibio.com). The ligation mix was transformed into DH10B through electroporation. High-density nylon filters (eight 384-well plates) were screened for a tiling path that covers the entire genome. Shotgun plasmid libraries were made from minimal tiling clones in pUC-18 vector.

Thermo-cycling sequencing reaction was performed in a final volume of 24 μL containing 16-μL DYEnamic ET Terminator sequencing kit premix, 10 pM universal sequencing primers, and 500 ng plasmid DNA. The reaction conditions were 95°C for 2 min, followed by 35 cycles of 95°C denaturation for 15 s, 50°C annealing for 15 s, and 60°C extension for 90 s. The amplified DNA fragments were sequenced on ABI-3730 DNA sequencer. DNA sequences were assembled by using the software package phred/phrap/consed/ (Ewing and Green 1998; Gordon *et al*. 1998) on a PC/UNIX platform. The mitochondrial sequences were annotated with glimmer 3.0 and BLAST tools, and tRNA genes and their secondary structures were identified according to tRNAscan-SE (Lowe and Eddy 1997).

Evolutionary analysis

Complete mitochondrial genome sequences used for comparative and evolutionary analyses were obtained from the NCBI database (http://www.ncbi.nlm.nih.gov/genome), and they are: *Triticum aestivum* cv. Chinese Yumai (GenBank, EU534409); *Marchantia polymorpha* (GenBank, M68929); *Physcomitrella patens* (GenBank, AB251495); *Beta vulgaris* subsp. vulgaris (GenBank, BA000009); *Brassica napus* (GenBank, AP006444); *Nicotiana tabacum* (GenBank, BA000042); *Arabidopsis thaliana* (GenBank, Y08501); *Oryza sativa* (GenBank, DQ167399); *Sorghum bicolor* (GenBank, DQ984518); *Tripsacum dactyloides* (GenBank, DQ984517); *T. aesitvum* (GenBank, AP008982) and *Zea mays* subsp. *mays* (GenBank, AY506529). Multiple alignments of protein-coding genes were performed with CLUSTAL W (Thompson *et al*. 1994) and used to estimate variation rates among synonymous (K_s) and nonsynonymous (K_a) sites with a maximum likelihood (ML) method (Yang and Nielsen 2000) implemented in K_a_K_s_Calculator (Zhang

et al. 2006). To test for positive selection, we used Fisher's exact test combined with the Nielsen–Yang method (Nielsen and Yang 1998), as implemented in $K_a_K_s$_Calculator (Zhang *et al*. 2006). This method tests for the presence of positively-selected sites based on K_a/K_s ratios that vary among sites. Rate constancy was tested with the Tajima's procedure applied in MEGA (Kumar *et al*. 2004).

Results

Genome organization

Since the sequences of plant mitochondrial protein-coding genes are highly homologous, we designed a set of PCR primers per probes for screening the wheat BAC library constructed from mtDNA (see table 1 in electronic supplementary material at http://www.ias.ac.in/jgenet/). We chose 44 clones that cover a minimal-tiling-path for shot-gun sequencing. We generated 9931 reads with an average length of 424 bp at a quality value of Q20; the sequences are collectively equivalent to nine-fold genome coverage.

Table 1. Nucleotide frequencies of one per codon position in higher plant mtDNA genes.

Codon position	A	T	G	C
Wheat				
1	0.095	0.08	0.089	0.07
2	0.081	0.113	0.063	0.075
3	0.091	0.115	0.064	0.063
Rice				
1	0.092	0.08	0.091	0.071
2	0.081	0.11	0.064	0.079
3	0.09	0.111	0.067	0.065
Arabidopsis				
1	0.093	0.08	0.086	0.075
2	0.084	0.109	0.064	0.077
3	0.087	0.107	0.07	0.07
Liverwort				
1	0.104	0.08	0.086	0.064
2	0.095	0.109	0.062	0.068
3	0.096	0.11	0.065	0.062

The complete mitochondrial genome of *Triticum aestivum* cv. Chinese Yumai is 452,526 bp in length and contains 33 known protein-coding genes, three rRNA and 17 tRNA genes as well as 149 ORFs greater than 300 bp in length (figure 1 and table 2 of electronic supplementary material). The sequence is almost identical to the previously reported sequence of *T. aestivum* cv. Chinese Spring (Ogihara *et al*. 2005), which further confirms the accuracy of the sequence reported previously. We identified seven SNPs (three transitions and four transversions) and 10 indels (insertions and deletions), all these were found in non-coding regions. A 4-bp indel serves as a convenient marker for discriminating the cultivar Chinese Yumai from Chinese Spring.

Nucleotide composition and codon usage

Among mtDNAs of wheat and other higher plant, nucleotide composition in protein-coding genes shows a common pattern (table 1). Nucleotide frequencies vary at different codon positions. At the first positions, A is used most frequently, as compared to the second and third codon positions where T is found most frequently. The relative frequencies of the four nucleotides for the three codon positions in order are A>G>T>C, T>A>C>G and T>A>G>C. Obviously, the AT-content, regardless of position, is always higher than the GC-content; this is consistent with the high AT-content (average 58%) of all plant mitochondrial genomes.

Mitochondrial genomes of higher plants share a common genetic code and similar codon usage. Codon usage patterns in wheat and other plant mtDNAs are summarized in table 2. For amino acids encoded by six or four codons, the codons ending with A or T are used more frequently than those ending with G or C. In addition, A-containing and G-containing, or purine-containing codons are used more frequently. Further, TTT for phenylalanine is the most frequently used codon whereas TGC for arginine is the least used. The nucleotide frequencies and codon usage patterns are shared by all higher plants, despite a long evolutionary process from non-flowering to flowering plants, and could be the result of a low evolutionary rate of plant mtDNAs (Ogihara *et al*. 2005; Tian *et al*. 2006).

An evolutionary analysis of protein-coding genes

We studied nucleotide and amino acid substitution rates among mitochondrial protein-coding genes among cereals (wheat, rice, maize and sorghum) through estimating the number of synonymous and nonsynonymous substitutions per synonymous (K_s) and nonsynonymous site (K_a), respectively. Nucleotide and amino acid substitution rates vary across mitochondrial genomes and within individual genes (see figure 1 in electronic supplementary material; figure 2A). Nucleotide substitution rates appeared to be lowest for the two families of cytochrome c biogenesis proteins (*ccmB, ccmC, ccmFC* and *ccmFN*) and the NADH ubiquinone dehydrogenase complex (*nad1, nad2, nad3, nad4, nad4L, nad5, nad7* and *nad9*) with an exception of *nad6* gene that has a relatively higher substitution rate. The three subunits of the cytochrome c oxidases (*cox1, cox2* and *cox3*) and the ATP synthases family (*atp1, atp4, atp8* and *atp9*) showed a relative higher substitution rate. The rates of ribosomal proteins vary widely, from the lower rates of *rps3, rps7*, and *rpl16* to the highest rate of the gene *rps2*.

The substitution rates among different gene families vary significantly across plant lineages. In contrast to monocotyledons, the rates of nucleotide substitution among mitochondrial genes of dicotyledons (rapeseed, tabacum, sugar beet and *Arabidopsis*) show different variable patterns. For instance, the lowest rates were found in genes for NADH

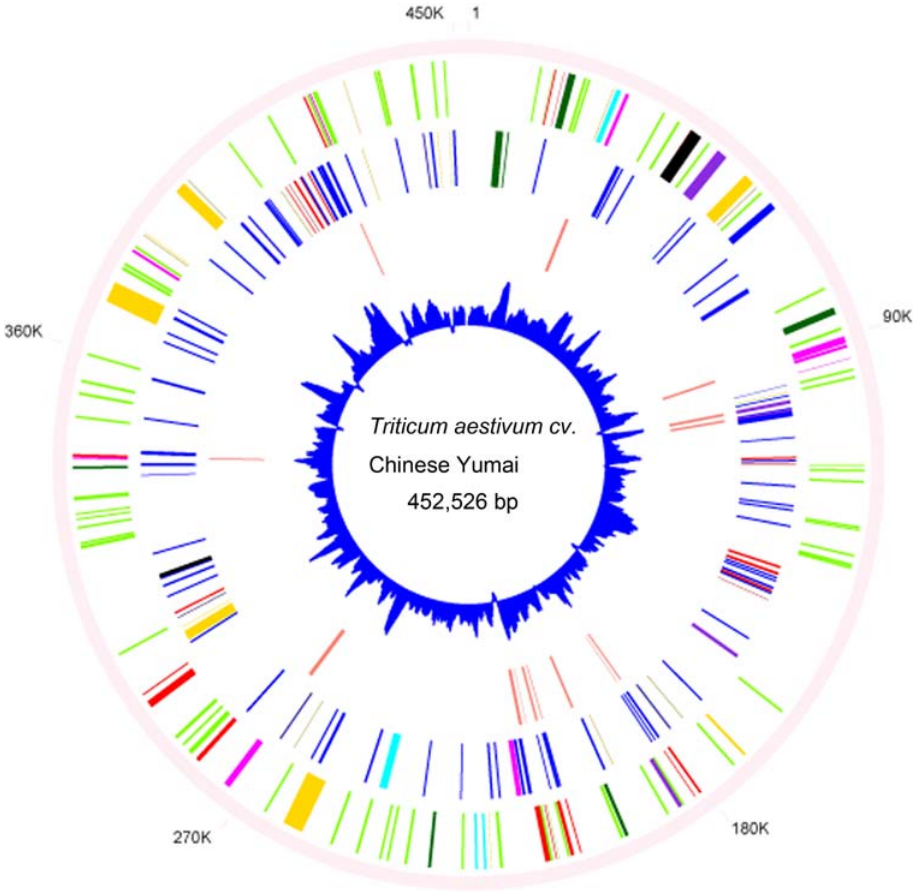

Figure 1. The physical map of the wheat (*Triticum aestivum* cv. Chinese Yumai) mitochondrial genome. Circles display (from outside): (1) physical map scaled in kilobase pairs, (2) and (3) coding sequences transcribed clockwise and counter clockwise, respectively: dark-green, ATP synthases; red, NADH ubiquinone dehydrogenases; blueviolet, cytochrome c biogenesis proteins; cyan, cytochrome c oxidases; blue, cytochrome b oxidase; fuchsia, ribosomal proteins; gold, tRNA and rRNA genes; black, maturase and mttB; chartreuse, open reading frames; (4) the genes shown accelerated rate of evolution; (5) GC content variations (in a 1000-bp window and 100-bp increments).

Table 2. Comparison of codon usage (number of codons) among wheat and selected plants.

Amino acid[a]	Codon	Wheat	Rice	*Arabidopsis*	Liverwort
A	GCA	174	258	343	300
	GCC	171	215	327	291
	GCG	105	156	193	200
	GCT	290*	398	526	454
C	TGC	66	106	169	156
	TGT	110*	136	250	225
D	GAC	116	162	276	289
	GAT	298*	374	499	448
E	GAA	333*	441	597	550
	GAG	169	226	412	317
F	TTC	313	404	615	384
	TTT	489*	580	826	947
G	GGA	294*	371	498	446
	GGC	118	157	220	252
	GGG	142	222	309	243

Table 2 (*contd*)

Amino acid[a]	Codon	Wheat	Rice	*Arabidopsis*	Liverwort
	GGT	271	301	421	419
H	CAC	71	96	177	175
	CAT	230*	270	393	324
I	ATA	273	329	445	547
	ATC	258	307	481	311
	ATT	406*	479	589	724
K	AAA	318*	389	610	972
	AAG	221	273	479	475
L	CTA	189	228	372	312
	CTC	152	195	391	195
	CTG	132	175	323	242
	CTT	273	357	566	421
	TTA	315*	360	525	687
	TTG	247	342	463	444
M	ATG	325	400	588	472
N	AAC	128	173	298	335
	AAT	292*	331	478	537
P	CCA	193	250	331	231
	CCC	141	199	272	203
	CCG	83	124	207	145
	CCT	209*	260	460	316
Q	CAA	267*	338	489	531
	CAG	90	134	278	212
R	AGA	201*	265	370	323
	AGG	103	147	283	224
	CGA	156	206	264	239
	CGC	72	96	164	158
	CGG	96	125	203	115
	CGT	158	192	252	247
S	AGC	121	148	266	233
	AGT	191	227	336	301
	TCA	197	266	358	280
	TCC	178	244	363	248
	TCG	139	180	258	183
	TCT	236*	298	519	330
T	ACA	149	182	314	272
	ACC	154	197	282	271
	ACG	95	112	181	193
	ACT	208*	278	399	361
V	GTA	199	239	356	324
	GTC	139	174	271	195
	GTG	163	205	288	275
	GTT	220*	285	391	407
W	TGG	175	242	348	323
Y	TAC	90	126	227	249
	TAT	283*	335	474	516

[a] The abbreviation of amino acids. *For amino acids encoded by multiple codons, the codons ending with A or T are used more frequently.

ubiquinone oxidoreductase complex and ATP synthases, whereas a relative higher rate was shown among other genes. The rates of amino acid substitutions appeared to have higher rates in *atp4, rps3, cox2*, and *ccmFN*, and all these genes do not have a higher rate in other plant lineages (see figure 1B in electronic supplementary material; figure 2B).

Among the early land plants, such as liverwort and moss, the nucleotide and amino acid substitution rates are quite different from those of flowering plants, being are lower for ribosomal proteins, ATP synthases, and cytochrome c biogenesis proteins but higher for the NADH ubiquinone oxidoreductase complex and cytochrome c oxidases, especially

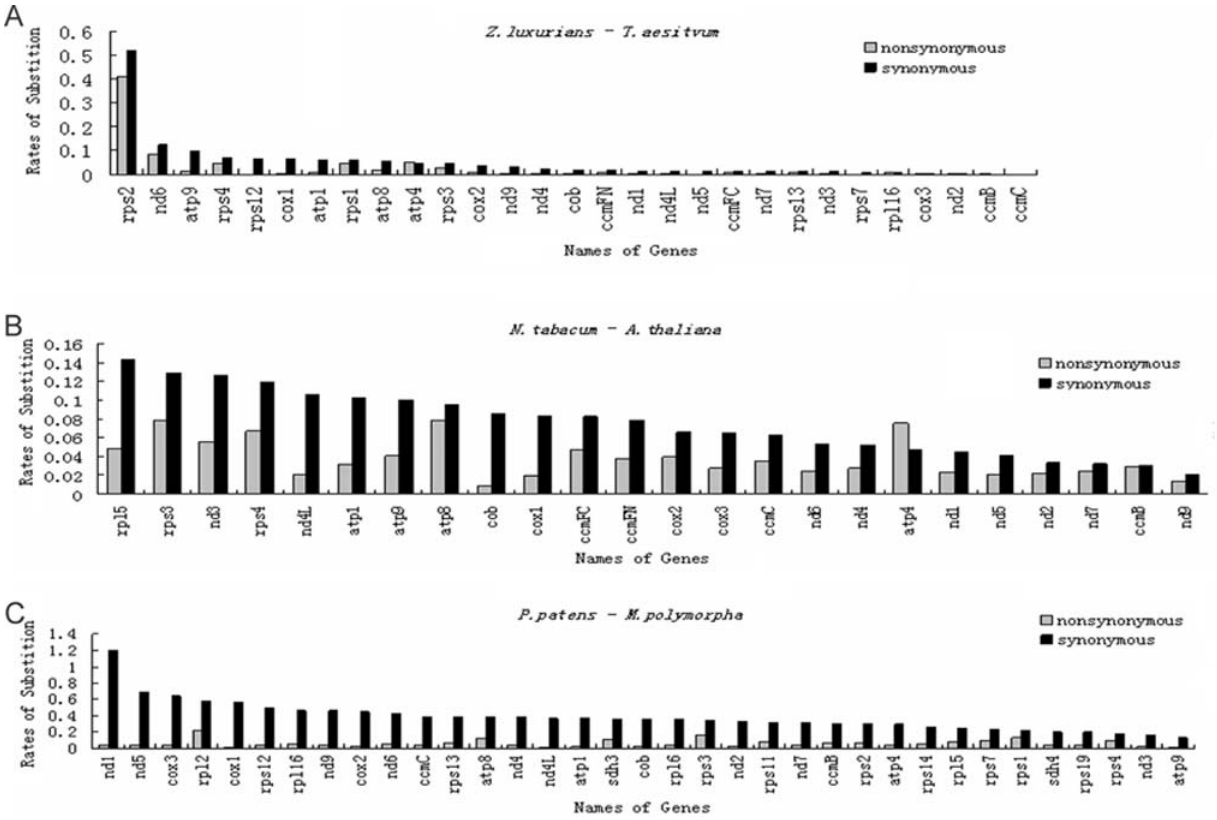

Figure 2. The nonsynonymous and synonymous substitution rates among othologous mitochondrial gene-pairs of selected higher plants. A, the rates varied among gene-pairs from monocotyledons (maize–wheat); B, the rates varied among gene-pairs from dicotyledons (tabaccco–*Arabidopsis*); C, the rates varied among gene-pairs from lower land plants (liverwort–moss). The genes of NADH dehydrogenase are abridged as nd.

in *nad1* that has the highest rate. In addition, the rates of amino acid substitution are higher in *ccmFN, ccmFC, rpl16* and *rps3* (see figure 1C in electronic supplementary material; figure 2C).

The fast evolution of mitochondrion-encoded genes among higher plants

We used relative rate tests to study evolutionary rates of mitochondrial genomes among plant lineages based on MEGs. The test statistic is asymptotically chi-square distributed and can be used to determine whether the number of nucleotide differences in two lineages is significantly different (Tajima 1993). The result demonstrated that evolutionary rates vary among plant lineages. The comparisons between dicotyledon and monocotyledon early showed significant differences (e.g., rice-sugar beet); moreover, many intra-group comparisons were significant (e.g., wheat–rice, tabacum–brisscia and liverwort–moss) (table 3). The different evolutionary rates could be correlated with substantial variations in metabolic rate and generation time among plant lineages. We used the maximum likelihood method to estimate the ratio (K_a/K_s) of the non-synonymous to synonymous substitution rate in higher plant mtDNA genes. Among monocotyledon mtDNAs (wheat, rice, maize, sorghum and *Tripsacum*), 30 genes were found to have orthologs. We first constructed 116 orthologous gene-pairs, and then estimated the K_a/K_s ratio between them (see table 3 in electronic supplementary material; figure 3). We found 55 gene-pairs in K_a/K_s ratio that is more than 0.5, 28 of which show K_a is greater than K_s (i.e., $K_a/K_s > 1$). Especially, in the genes *ccmB, ccmC, nad2, nad3, nad4L*, and *nad9*, the K_a/K_s ratio appears to be infinite, due to the rates of synonymous substitutions equal-

Table 3. Relative rate test among pairs of plant lineages.

Taxon pairs	Outgroup	χ^2	P value
Wheat Rice	Sorghum	233.860	0.000
Maize Sorghum	Wheat	2.630	0.100
Brassica *Arabidopsis*	Tabacum	1.680	0.194
Tabacum Brassica	Sugar beet	8.860	0.003
Wheat *Arabidopsis*	Liverwort	1.440	0.230
Rice Sugar beet	Moss	55.740	0.000
Moss Liverwort	*Arabidopsis*	10.620	0.001

ing to zero. These genes could suffer from the back mutation, but present method of calculating K_a/K_s cannot accurately detect it. Similarly, in dicotyledons (*Arabidopsis*, rapeseed, sugar beet and tabacum), 72 orthologous gene-pairs have been studied. Thirty-three genes show higher K_a/K_s ratio ($K_a/K_s > 0.5$), eight of which show $K_a/K_s > 1$. Totally,

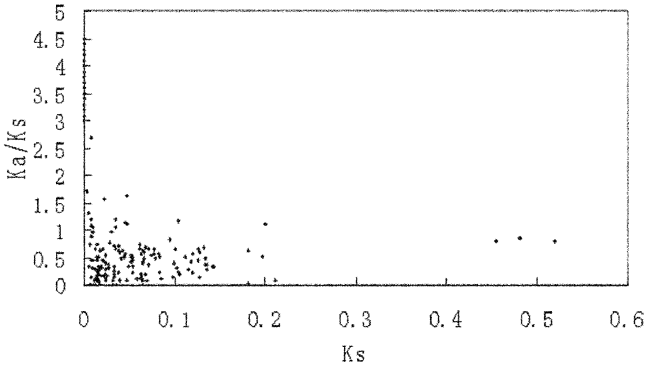

Figure 3. The distribution of K_a/K_s ratios among all orthologous gene-pairs of selected plant mitochondrial genomes. The dots on the vertical axis indicate that K_a/K_s ratios are infinite.

36 gene-pairs show a K_a/K_s ratio of more than one, which are from five gene families involved in energy production. For these gene-pairs, we performed a statistical test for positive selection, using Fisher exact test with P value (table 4); the result shows no significant evidence for positive selection ($P > 0.05$).

Discussion

In this study, we determined the mitochondrial genome sequence of *T. aestivum* cv. Chinese Yumai. Based on comparative analysis with those of other higher plants, we found a remarkable pattern that half of the orthologous gene-pairs from seed plants show higher K_a/K_s ratio, which means that the accelerated rate of evolution occurred in angiosperm mtDNA genes. It is generally believed that the accelerated rate of evolution could be due to positive selection or relaxation of purifying selection (Wyckoff *et al*. 2000; Arbiza *et al*. 2006).

Under the positive selection hypothesis, K_a/K_s can be high if advantageous replacement mutations are often fixed in the population. The action of positive selection is almost certain if the K_a/K_s ratio is significantly greater than one (Wyckoff *et al*. 2000; Arbiza *et al*. 2006). We performed the test for positive selection; however, the test showed no significant result. We thus believe that positive selection is not a major factor for the accelerated rate of evolution in the seed plant mtDNA. However, we cannot completely exclude the possibility that positive selection has acted on these genes. The K_a/K_s ratio is an average over all sites, and even under the action of positive selection, it can be smaller than one because some sites might be under positive selection, whereas others are under purifying selection (Wernegreen and Riley 1999; Betran *et al*. 2006).

Under the relaxation hypothesis, if the protein does not have a strong functional constraint, rate of protein evolution would be accelerated, and K_a/K_s would be high (Arbiza *et al*. 2006; Betran *et al*. 2006). The relaxation of selection could play an important role in the shape of this evolutionary pattern in seed plants. Mitochondria provide the primary source of cellular ATP in eukaryotes via the process of oxidative phosphorylation (Lee and Wei 2000). In plants, mitochondrial genomes mainly code for 22 subunits of oxidative phosphorylation and several ribosomal proteins. The selection relaxation acting on these genes enhance rate of protein evolution, and at the same time, promoted the divergence of proteins function among species.

We believe that the relaxation of selection not only act on the sequences of protein-coding genes, but also could act on the genome structure. Higher plant mitochondrial genomes

Table 4. Genes that show $K_aK_s > 1$.

Genes	Function	Nubmer of gene-pairs	Fisher's test P-value
atp4	ATP synthases	5	0.2525
ccmB	Subunit of cytochrome c biogenesis proteins	5	0.8437
ccmC	Subunit of cytochrome c biogenesis proteins	3	–
ccmFC	Subunit of cytochrome c biogenesis proteins	2	0.9866
cox2	Subunit of cytochrome c oxidases	1	–
cox3	Subunit of cytochrome c oxidases	4	0.8473
nad2	Subunit of NADH ubiquinone oxidoreductase complex	4	0.8914
nad3	Subunit of NADH ubiquinone oxidoreductase complex	2	–
nad4L	Subunit of NADH ubiquinone oxidoreductase complex	3	–
nad7	Subunit of NADH ubiquinone oxidoreductase complex	1	–
nad9	Subunit of NADH ubiquinone oxidoreductase complex	1	–
rpl16	The large subunit of ribosomal proteins	2	0.8284
rps1	The small subunit of ribosomal proteins	1	0.5484
rps4	The small subunit of ribosomal proteins	2	0.6847

–, There is no test result because of the othologous genes show $K_s = 0$.

appear to be highly variable in structure, including genome size and gene content. The size of higher plant mitochondrial genomes varies widely, ranging from 200 to 800 kb, largely due to the expansion of intergenic region, such as insertions of non-mitochondrial sequences from nuclear, chloroplast, and viral DNAs as well as their duplicated sequences. In addition, some genes, such as ribosomal proteins, rRNA and tRNA genes, have been lost during the process of higher plant evolution; further, the gene of *atp6*, the subunit of ATP synthases, show a significant change in length and amino acids sequences (Kubo *et al*. 1993; Handa 2003; Clifton *et al*. 2004; Ogihara *et al*. 2005; Tian *et al*. 2006). In our analysis, the gene of *atp6* was eliminated because of suspected alignment errors and frame-shift mutation. Above all, relaxation of selective constraint on seed plant mtDNA could accelerate the evolution of the sequences and structures of the genomes.

In contrast to plant mtDNAs, animal mitochondrial genomes appear to be more conserved in size (14–19 kb) and gene contents (13 protein-coding genes, 22 tRNA, and two rRNA genes (Gray *et al*. 1999)). Their protein-coding genes exhibit a higher synonymous substitution rate than that of nonsynonymous substitutions between species (Ballard and Kreitman 1994; Hasegawa *et al*. 1998; Ballard and Whitlock 2004; Bjornerfeldt *et al*. 2006). The test of selection has suggested that the evolution of animal mtDNAs is governed by many selective forces, and it deviates from a strictly neutral model (Ballard and Whitlock 2004; Bazin *et al*. 2006).

Acknowledgements

This work was supported by National Natural Science Foundation of China (30471081).

References

Arbiza L., Dopazo J. and Dopazo H. 2006 Positive selection, relaxation, and acceleration in the evolution of the human and chimp genome. *PLoS Comput. Biol.* **2**, e38.

Ballard J. W. and Kreitman M. 1994 Unraveling selection in the mitochondrial genome of *Drosophila. Genetics* **138**, 757–772.

Ballard J. W. and Whitlock M. C. 2004 The incomplete natural history of mitochondria. *Mol. Ecol.* **13**, 729–744.

Bazin E., Glemin S. and Galtier N. 2006 Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**, 570–572.

Betran E., Bai Y. and Motiwale M. 2006 Fast protein evolution and germ line expression of a *Drosophila* parental gene and its young retroposed paralog. *Mol. Biol. Evol.* **23**, 2191–2202.

Bjornerfeldt S., Webster M. T. and Vila C. 2006 Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res.* **16**, 990–994.

Chaw S. M., Chun-Chieh Shih A., Wu Y. W., Liu S. M., Chou T. Y. and Wang D. 2008 The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, *Bpu* sequences, and abundant RNA editing sites. *Mol. Biol. Evol.* **25**, 603–615.

Clifton S. W., Minx P., Fauron C. M., Gibson M., Allen J. O., Sun H. *et al*. 2004 Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol.* **136**, 3486–3503.

Ewing B. and Green P. 1998 Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194.

Gordon D., Abajian C. and Green P. 1998 Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202.

Gray M. W., Burger G. and Lang B. F. 1999 Mitochondrial evolution. *Science* **283**, 1476–1481.

Handa H. 2003 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.* **31**, 5907–5916.

Hasegawa M., Cao Y. and Yang Z. 1998 Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species. *Mol. Biol. Evol.* **15**, 1499–1505.

Itoh T., Martin W. and Nei M. 2002 Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc. Natl. Acad. Sci. USA* **99**, 12944–12948.

Kubo T., Mikami T. and Kinoshita T. 1993 The sugar beet mitochondrial genome contains an ORF sharing sequence homology with the gene for the 30 kDa subunit of bovine mitochondrial complex I. *Mol. Gen. Genet.* **241**, 479–481.

Kubo T., Nishizawa S., Sugawara A., Itchoda N., Estiati A. and Mikami T. 2000 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*. **28**, 2571–2576.

Kumar S., Tamura K. and Nei M. 2004 MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* **5**, 150–163.

Lee H. C. and Wei Y. H. 2000 Mitochondrial role in life and death of the cell. *J. Biomed. Sci.* **7**, 2–15.

Lowe T. M. and Eddy S. R. 1997 tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964.

Lynch M., Koskella B. and Schaack S. 2006 Mutation pressure and the evolution of organelle genomic architecture. *Science* **311**, 1727–1730.

Nielsen R. and Yang Z. 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929–936.

Notsu Y., Masood S., Nishikawa T., Kubo N., Akiduki G., Nakazono M. *et al*. 2002 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* **268**, 434–445.

Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N. *et al*. 1992 Gene organization deduced from the complete sequence of liverwort Marchantia polymorpha mitochondrial DNA. A primitive form of plant mitochondrial genome. *J. Mol. Biol.* **223**, 1–7.

Ogihara Y., Yamazaki Y., Murai K., Kanno A., Terachi T., Shiina T. *et al*. 2005 Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Res.* **33**, 6235–6250.

Osoegawa K. and de Jong P. J. 2004 BAC library construction. *Methods Mol. Biol.* **255**, 1–46.

Sugiyama Y., Watase Y., Nagase M., Makita N., Yagura S., Hirai A. and Sugiura M. 2005 The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Mol. Genet. Genomics* **272**, 603–615.

Tajima F. 1993 Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**, 599–607.

Terasawa K., Odahara M., Kabeya Y., Kikugawa T., Sekine Y., Fujiwara M. and Sato N. 2007 The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants. *Mol. Biol. Evol.* **24**, 699–709.

Thompson J. D., Higgins D. G. and Gibson T. J. 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

Tian X., Zheng J., Hu S. and Yu J. 2006 The rice mitochondrial genomes and their variations. *Plant Physiol.* **140**, 401–410.

Wernegreen J. J. and Riley M. A. 1999 Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: a case for the genetic coherence of rhizobial lineages. *Mol. Biol. Evol.* **16**, 98–113.

Wyckoff G. J., Wang W. and Wu C. I. 2000 Rapid evolution of male reproductive genes in the descent of man. *Nature* **403**, 304–309.

Yang Z. and Nielsen R. 2000 Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**, 32–43.

Zhang Z., Li J. and Yu J. 2006 Computing K_a and K_s with a consideration of unequal transitional substitutions. *BMC Evol. Biol.* **6**, 44.

Received 11 November 2008, in revised form 14 June 2009; accepted 21 July 2009

Published on the Web: 3 November 2009