SCIENTIFIC REPORTS

OPEN

Received: 2 November 2017
Accepted: 27 July 2018
Published online: 24 August 2018

Complete sequence of kenaf (*Hibiscus cannabinus*) mitochondrial genome and comparative analysis with the mitochondrial genomes of other plants

Xiaofang Liao[1,2,3], Yanhong Zhao[3], Xiangjun Kong[2], Aziz Khan[2], Bujin Zhou[2], Dongmei Liu[4], Muhammad Haneef Kashif[2], Peng Chen[2], Hong Wang[5] & Ruiyang Zhou[2]

Plant mitochondrial (mt) genomes are species specific due to the vast of foreign DNA migration and frequent recombination of repeated sequences. Sequencing of the mt genome of kenaf (*Hibiscus cannabinus*) is essential for elucidating its evolutionary characteristics. In the present study, single-molecule real-time sequencing technology (SMRT) was used to sequence the complete mt genome of kenaf. Results showed that the complete kenaf mt genome was 569,915 bp long and consisted of 62 genes, including 36 protein-coding, 3 rRNA and 23 tRNA genes. Twenty-five introns were found among nine of the 36 protein-coding genes, and five introns were *trans*-spliced. A comparative analysis with other plant mt genomes showed that four syntenic gene clusters were conserved in all plant mtDNAs. Fifteen chloroplast-derived fragments were strongly associated with mt genes, including the intact sequences of the chloroplast genes *psaA*, *ndhB* and *rps7*. According to the plant mt genome evolution analysis, some ribosomal protein genes and succinate dehydrogenase genes were frequently lost during the evolution of angiosperms. Our data suggest that the kenaf mt genome retained evolutionarily conserved characteristics. Overall, the complete sequencing of the kenaf mt genome provides additional information and enhances our better understanding of mt genomic evolution across angiosperms.

Mitochondria are the main organelles responsible for plant energy metabolism and play an imperative role in supplying ATP via oxidative phosphorylation during development, reproduction and various biochemical processes in plants. According to endosymbiotic theory, plant mitochondria are thought to be descended from free-living bacteria, which explains the presence of their genomes[1]. The structure of the plant mitochondrial (mt) genome has undergone dramatic changes over long-term evolution. Horizontal transfer with frequent exchanges among the nucleus, plastids and mitochondria appears to be responsible for the acquisition of exogenous sequences[2]. In addition, the abundance of repeated sequences of various sizes and numbers is involved in mt genome homogeneous recombination[3]. Thus, the noncoding regions vary and exhibit low conservation across species, which renders the sequencing of plant mt genomes, particularly in angiosperms, extraordinarily difficult. The first report of an angiosperm mt genome was achieved in *Arabidopsis thaliana*[4]. With recent sequencing efforts over the past decade, the mitochondria of many angiosperm species (e.g., *Beta vulgaris*[5], *Oryza sativa*[6], *Brassica napus*[7], *Zea mays*[8], *Triticum aestivum*[9], *Nicotiana tabacum*[10], *Vitis vinifera*[11], *Citrullus lanatus*[12], *Vigna radiata*[3], *Cucumis melo*[13], *Gossypium hirsutum*[14,15] and other higher plants[16–19]) have been sequenced. DNA sequencing and physical

[1]College of Life Sciences and Technology, Guangxi University, Nanning, 530005, China. [2]Key Laboratory of Plant Genetic and Breeding, College of Agriculture, Guangxi University, Nanning, 530005, China. [3]Cash Crop Institute of Guangxi Academy of Agricultural Sciences, Nanning, 530007, China. [4]Key Laboratory of Plant-Microbe Interactions, Department of Life Science and Food, Shangqiu Normal University, Shangqiu, 476000, China. [5]Department of Biochemistry, University of Saskatchewan, Saskatoon, SK, S7N5E5, Canada. Correspondence and requests for materials should be addressed to R.Z. (email: ruiyangzhou@aliyun.com)

mapping have been used to identify several evolutionarily conserved properties of plant mt genomes, i.e., gene order, genome structure, and migration of sequences from other organelles.

Angiosperm mt genomes are complex and vary substantially in size, ranging from 208 kb in *Brassica hirta*[5] to 11.3 Mb in *Silene conica*[20]. Despite the great variation in size and physical mapping properties, plant mitochondria exhibit significant conservation in functional genes, including 37–83 protein coding, tRNA and rRNA genes[21]. The shuffling of mtDNA sequences by recombination, repeat sequences and most noncoding sequences plays an important role in mt genome evolution by changing the gene organization and creating chimeric genes[22,23]. In most plant mt genomes, many homologous sequences are derived from the chloroplasts and nucleus[6,9]. In *Cucumis melo* mt genomes, 35 DNA fragments were found to originate from the chloroplast genome, while 1,114 DNA fragments with a total length of 1,272.6 kb were homologous with the nuclear genome, accounting for 46.5% of the mt genome[13]. Furthermore, horizontal gene (or DNA) transfers appear to be responsible for the integration of exogenous DNA and explain the complex structure of angiosperm mt genomes[24,25].

Kenaf (*Hibiscus cannabinus*) is an important fibre crop that is widely used in paper-making and weaving[26]. However, data regarding the mt genome sequence of kenaf are limited. Here we report the first complete kenaf mt genome of UG93B, which was a maintainer line and derived from the wild type of UG93. In the present study, the structure of first the complete kenaf mt genome sequence was determined, and phylogenetic analyses were performed for comparisons with angiosperm mt genomes. Our data provide basic information and a better understanding of the evolutionary processes of kenaf mt genome.

Results

Kenaf mitochondrial genome sequencing and assembly.

Isolated kenaf mitochondrial DNA (mtDNA) was used to construct a library for sequencing using PacBio RS II single-molecule real-time sequencing technology (SMRT), which generated 1.12 G of raw data, with an average read length of 4.6 kb, and the longest read was 32 kb. In total 67,152 reads (363,717,023 bp) were obtained after removing the adapter and low-quality regions, and the average coverage depth, read length and read quality were 605×, 5.4 kb, and 0.81, respectively (Supplementary Table S1). In total 1,819 reads (12,114, 267 bp, average length of 6.7 kb) were obtained after correcting by mapping the short reads to the long seed reads. After filtering the chloroplast reads, 1,762 reads (11,733,852 bp, average length of 6.7 kb) were used for the assembly process. Finally, the kenaf mt genome was assembled into a single circular molecule with a total length of 569,915 bp and an overall GC content of 44.9% (Fig. 1, Supplementary Table S2).

Gene content in kenaf.

Sixty-two genes including 36 core protein-coding genes, conserved among all plant mt genomes were annotated by comparing the assembled kenaf mt sequence with known plant mt sequences in the NCBI public DNA database using BLASTn. The kenaf mt genome contains 20, 7, 4, 4, 3, 29, 1 and 1 genes responsible for electron transport, oxidative phosphorylation, small ribosomal proteins, large ribosomal proteins, cytochrome C maturation protein, rRNAs, tRNAs, and *matR* and *mttB*, respectively (Supplementary Table S3). Most protein-coding genes, except for *sdh3*, *rps13* and *rps19*, were identical to those in the mt genome of the *Gossypium* species (Fig. 2). Twenty-three tRNA genes specifying 18 amino acids were identified in the kenaf mt genome. Of these genes, 15 tRNAs had a mt origin, and eight tRNA had a chloroplast origin (Table S3). The presence and locations of these genes in the kenaf mt genome and comparisons with other plant mt genomes are shown in Fig. 1 and Supplementary Table S4.

In most spermatophytes, the genes responsible for the electron transport chain and oxidative phosphorylation are conserved, except for mitochondrial complex II, which contains *sdh3* and *sdh4*. Notably, the high diversity in the gene content among the higher plant mt genomes was a primary contributor to the variety of ribosomal protein genes (Fig. 2). The mt genome of plants is known to contain genes encoding products involved in electron transport, oxidative phosphorylation, ATP synthesis, cytochrome c biogenesis, ribosomes, and the translation of proteins(Fig. 2).

The protein-coding genes in the kenaf mt genome account for 6.9% of the genome and a total length of 39,534 bp. In addition, 126 open reading frames (ORFs) larger than 100 amino-acid residues in size were annotated in the kenaf mt genome (Supplementary Table S2). However, none of these ORFs could be assigned a function based on sequence similarity at either the nucleotide or protein level. Most of ORFs were considered hypothetical proteins. A putative protein of 295 amino-acid residues encoded by ORF295 has unknown functions, although its first 183 nucleotides were similar to those of *rps4* from the 5′ to 3′ end (Supplementary Fig. S2). The nucleotide sequence in the coding region had no similarity to any other plant mt genomes, except for a part of the *Gossypium* mt genome. This chimeric characteristic of ORF295 may have resulted from a horizontal gene transfer (HGT) event between angiosperm mt genomes. Interestingly, despite the large size differences among the mt genomes of various higher plant species, these genomes share a similar set of functional genes (protein, rRNA and tRNA genes), which is consistent with the results reported by Mower *et al.*[27]. However, the additional ORF e.g., ORFs identified in the kenaf mt genome, were not shared even among closely related plants, suggesting that many ORFs likely do not encode functional proteins and may have unidentified species-specific functions.

Repeat sequences of kenaf mitochondrial DNA.

Repeat sequences are extensively found in the plant mt genome, are characterized primarily as forward repeat and palindromic repeats, and exhibit high levels of polymorphism. In the present study, we identified 584 repeat sequences that ranged from 20 to 7,782 bp and accounted for 11.71% of the total kenaf mt genome (Supplementary Table S2, Fig. S3). Most repeats (approximately 95%) were between 20 and 100 bp in length, accounting for 6.63% of the total genome; approximately 5% (28) of the repeats were larger than 100 bp, and three repeats were larger than 1 kb (R1, 7,782 bp; R2, 1, 877 bp; and R3 1,528 bp) (Table 1). Most repeat sequences (≥60 bp) contained 2 copies of the repeat, and eight repeat sequences contained three copies (Table 1).

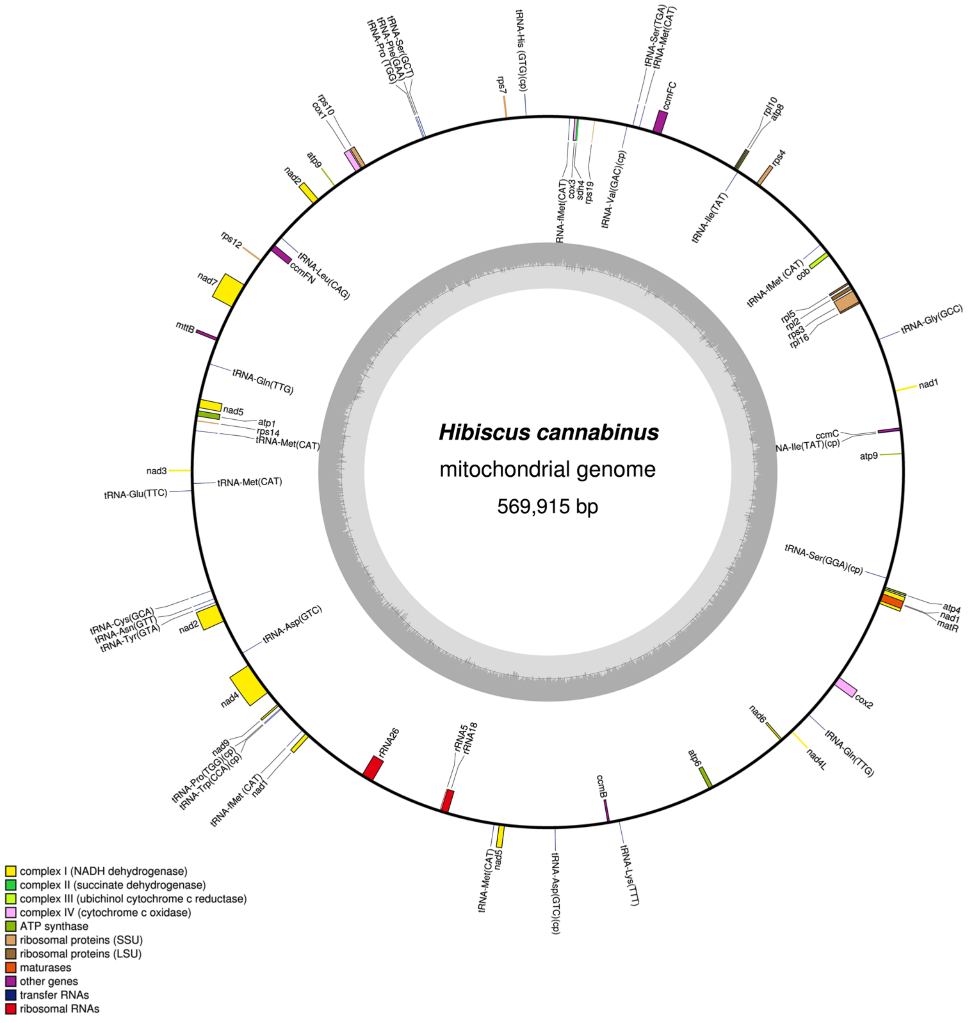


Figure 1. Map of the *Hibiscus cannabinus* (kenaf) mt genome.

Introns.

In the kenaf mt genome, nine mt genes, composed of 25 introns ranging from 41 to 2,878 bp in size were identified, and occupied 7.8% of the total kenaf mt genome (Supplementary Table S4). Five of the nine mt genes were *nad1, nad2, nad4, nad5* and *nad7*, which are components of mitochondrial complex I, and the remaining four genes were *cox2*, *ccmFc*, *rps3* and *rps10*. In addition, five *trans*-spliced introns observed in *nad1, nad2* and *nad5* were fragmented into separate coding regions, which is consistent with angiosperm plants. Twenty *cis*-spliced intron sequences were observed in the remaining mt genes. The intron locations and splicing were highly similar to those observed in other higher plant mt genomes (Fig. 2).

Chloroplast-like sequences.

BLASTn was used to identify chloroplast-like sequences in the kenaf mt genome. Twelve such sequence fragments were identified and showed >97% nucleotide sequence identity with the corresponding chloroplast sequences, and the segments ranged from 73 bp to 2,653 bp with a total length of 11,281 bp (accounting for 1.98% of the genome size). These chloroplast-derived fragments included eight tRNA-related sequences. Moreover, three intact chloroplast-related genes, i.e., *ndhB*, *psaA* and *rps7*, were identified in the kenaf mt genome (Supplementary Table S5).

Gene organization and gene clusters in plant mt genomes.

The gene organization greatly differs among plant mt genomes. In this study, we compared the gene orders in the 28 mt genomes and counted the number of syntenic gene clusters (genes that remain in the same order). Four gene clusters (i.e., *rrn5-rrn18, nad1-matR*, *rps12-nad3*, and *rps3-rpl16*) were found to be highly conserved in the plant mt genomes (Fig. 3, Supplementary Table S6). The gene cluster *cox3-sdh4* was widely distributed in most dicotyledonous species, except for *Brassicaceae*, while the conserved *rpl5-rps14* gene cluster was scattered in the other dicotyledonous species, but present in all *Brassica* spp. Understandably, species that have close evolutionary relationships share more clusters. Each gene cluster is transcribed from the same strand, implying that the genes may undergo co-transcription as a polycistronic mRNA.

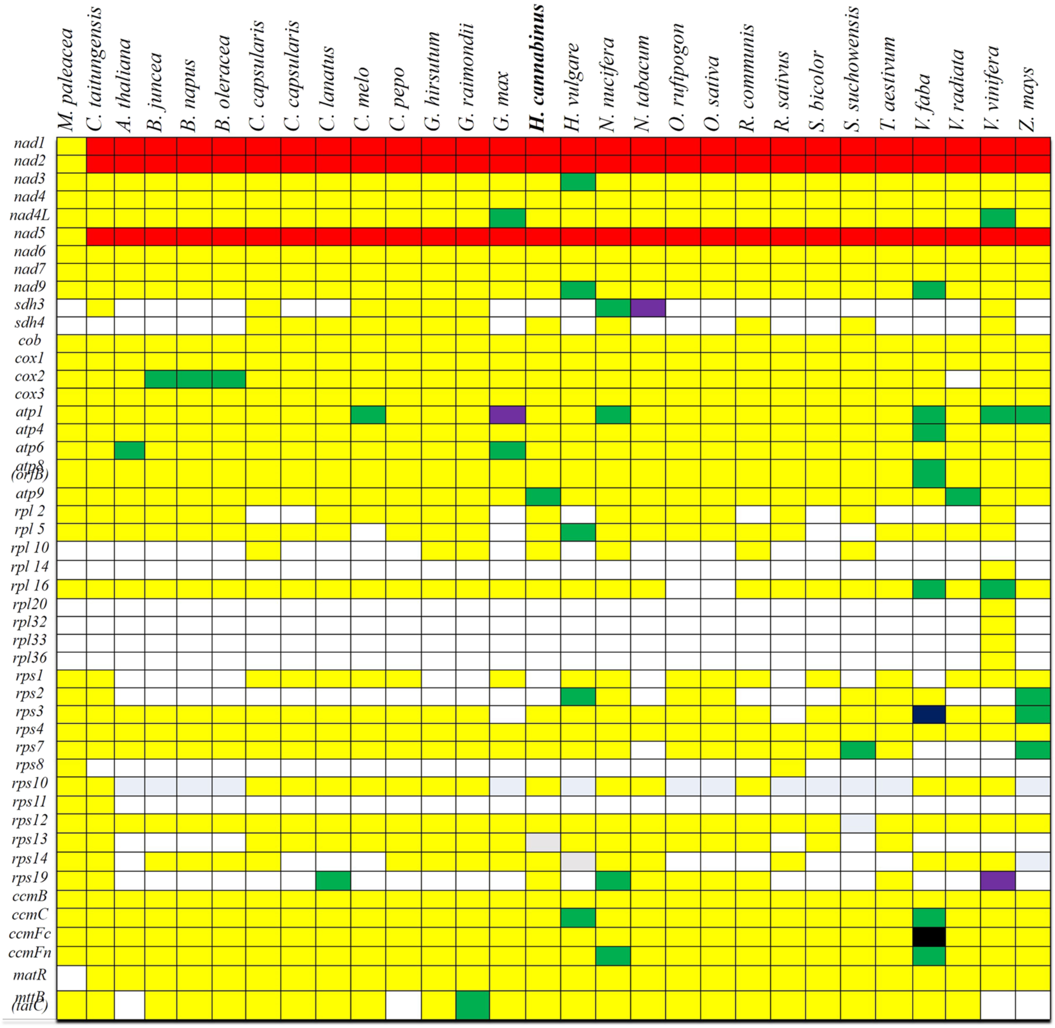


Figure 2. Distribution of protein-coding genes in plant mitochondrial genomes. White boxes indicate that the gene is not present in the mt genome. Yellow, green, purple, blue and black boxes indicate that one, two, three, four and six copies exist in the particular mt genome, respectively. Red boxes indicate trans-splicing. Kenaf (*Hibiscus cannabinus*) is shown in bold.

Distribution of tRNAs and DNA transfer from the plastid to mitochondrial DNA. A complete set of tRNAs is essential for protein translation in the plant mt genome. However, many tRNAs undergo loss, migration and inactivation during mt genome evolution in higher plants[27]. To evaluate the origin and distribution of the tRNA genes, tRNA scan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/) was used to predict the number and types of tRNA genes in the kenaf mt genome. In total, 23 tRNA genes were identified, and these genes recognized 18 amino acids (i.e., Asp, Gly, Met, Ser, His, Phe, Pro, Glu, Cys, Asn, Tyr, Trp, Asp, Lys, Ser, Leu, Ile, and Val). Thus, tRNA genes for two amino acids (i.e., Ala and Thr) were not identified and appeared to be missing from the kenaf mt genome (Fig. 4). Of these 23 tRNAs, eight had a plastid origin, and twenty-one had a mt origin.

The mt genomes of twenty-eight land plants and the fungal species *G. lucidum* were analysed to explore the patterns of tRNA loss during the evolution of plant mt genomes. Only *G. lucidum* has a complete set of tRNAs (Fig. 4). The *trnA* gene was lost from gymnosperms to angiosperms, indicating that *trnA* was lost early in the evolution of land plants. The *trnG* gene was absent from monocots, but existed in dicotyledons, suggesting that this gene was specifically present in dicotyledons. Although *trnL, trnR, trnT* and *trnV* were lost during the evolution of angiosperms, *trnR* and *trnV* existed in certain dicotyledons, suggesting that these genes may have been subsequently regained. Interestingly, most of the tRNAs in *C. melo* exhibited a pattern of plastid-like origin, suggesting that frequent exchanges occurred between the mt genome and the chloroplast genome.

BLASTn was used to assess the mt sequence fragments that originated in the chloroplast. Four chloroplast-derived fragments (*trnH, trnM, trnN* and *trnW*) were found to be conserved in all analysed mt genomes, and one (*trnD*) and two (*trnC* and *trnF*) chloroplast-derived fragments were found to be conserved in dicots and monocots, respectively. In contrast, other chloroplast-like tRNA genes exhibited scattered distributions, and certain native tRNA genes were irregularly lost among the higher plant mt genomes, suggesting that the gain and loss events of the tRNA genes occurred multiple times during evolution. Overall, *trnC, trnE, trnK, trnM,*

	Size	Identity	Copy-1		Copy-2[a]		Copy-3[a]		
No.	(bp)	(%)	Start	End	Start	End	Start	End	Type[b]
R1	7782	100	814	8595	196628	188847			P
R2	1877	100	62808	64684	356837	354961			P
R3	1525	100	60919	62443	446303	444779			P
R4	842	100	204427	205268	367743	384565			F
R5	535	100	62438	63272	445775	444941			P
R6	468	100	287766	288233	413387	412920			P
R7	433	100	62845	63277	136700	137132	358244	357812	F/P
R8	394	100	431948	432341	569460	569853			F
R9	374	100	4933	5306	199917	199544	380908	381281	P/F
R10	229	100	237021	237249	320344	320572			F
R11	210	100	28517	28726	243614	243405			P
R12	204	100	84633	84836	413895	414098			F
R13	190	100	326	515	204717	204528	368033	367844	P
R14	181	100	284825	284645	539224	539404			F
R15	174	100	73673	73846	539296	539469			F
R16	165	100	358549	358713	445775	445939			F
R17	146	100	247536	247681	510485	510630			F
R18	137	100	102129	102265	458421	458557			F
R19	135	100	683	817	204407	204273			P
R20	128	100	136700	136827	445775	445648			P
R21	115	100	683	797	367743	367629			P
R22	109	100	45409	45517	456623	456731			F
R23	109	100	73673	73781	284897	285005			F
R24	107	100	85072	85178	273126	273232			F
R25	103	100	176182	17714	380789	380891			F
R26	97	100	47563	47659	538666	538762			F
R27	91	100	362171	362261	532778	532868			P
R28	88	100	34694	34781	258881	258968			F
R29	88	100	102046	102133	137837	137924			P
R30	88	100	140828	140915	465024	465111			P
R31	86	100	62665	62750	268392	268477	445997	446082	F/ P
R32	82	100	31305	31386	301188	301269			F
R33	81	100	368575	368655	456163	456243			F
R34	80	100	4243	4322	200901	200980	268607	268686	P /F
R35	77	100	28446	28522	243819	243895			P
R36	75	100	333330	333404	429353	429427			P
R37	74	100	11547	11620	134973	135046			F
R38	73	100	41569	41641	438195	438267			F
R39	73	100	268577	268647	539445	539517			F
R40	70	100	189	258	204978	205047	368294	368373	P
R41	66	100	141384	141449	452383	452448			P
R42	65	100	4873	4937	200286	200350	380849	380913	P/F
R43	65	100	200286		380852				P
R44	63	100	141511	141573	452513	452575			F
R45	63	100	41704	41766	438338	438400			F
R46	61	100	47590	475650	275165	275225	538696	538756	P
R47	61	100	432343	432403	569854	569914			F
R48	61	100	73792	73852	285008	285068			F

Table 1. Repeats (≥60 bp) in the kenaf mt genome. [a]Compare with copy-1 as control. [b]The letters F and P represent forward repeats and palindromic repeats, respectively. The numbers listed in the starting and ending points refer to positions in the kenaf mt genome sequence (GenBank accession MF163174).

trnP, trnQ, trnS and *trnY* were present in all species evaluated, indicating that these tRNAs are highly conserved in plant mt genomes.

Conserved sequences and phylogenetic analysis. A phylogenetic analysis was performed to determine the evolutionary relationships among the mt genomes of twenty-eight plant species, included angiosperms

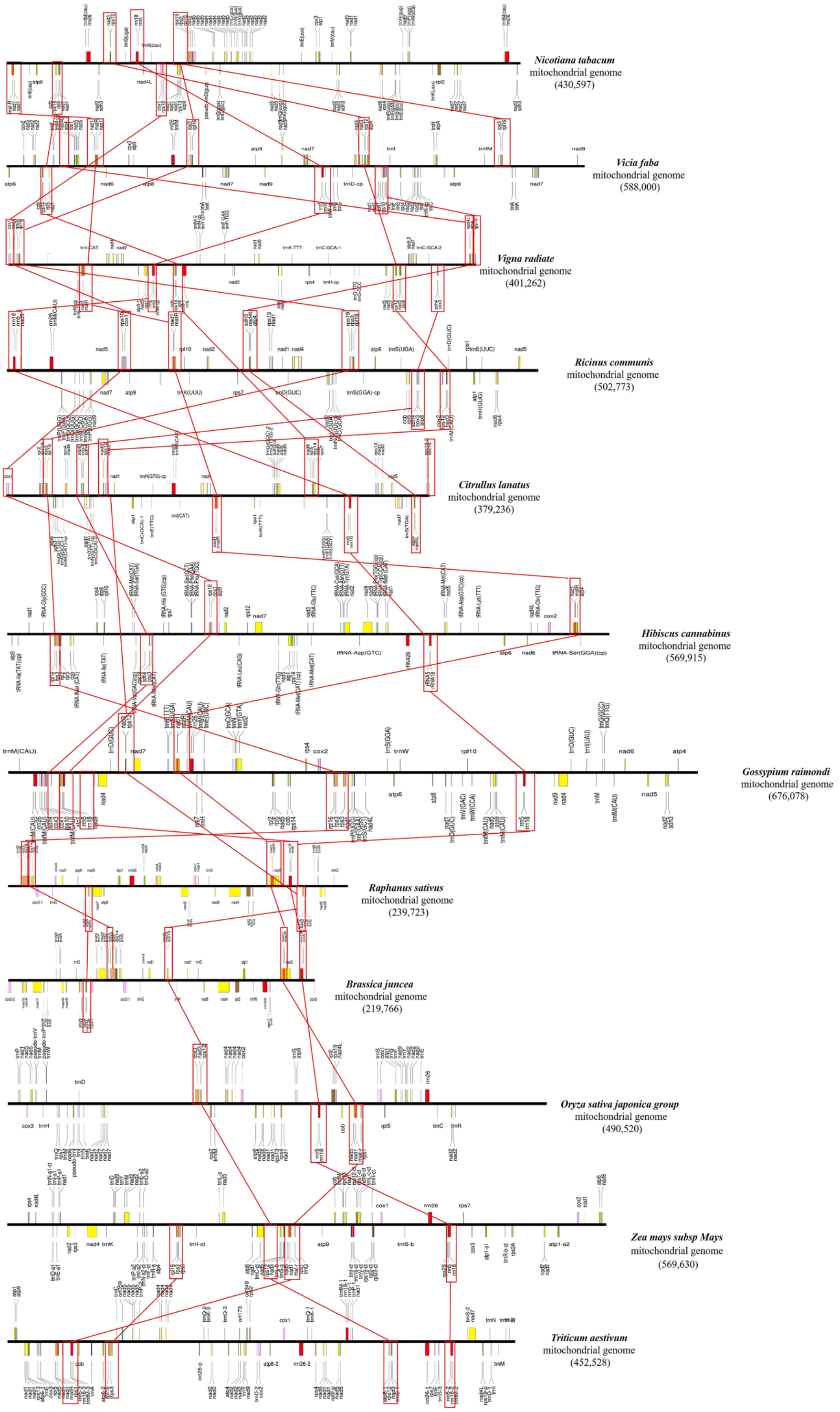
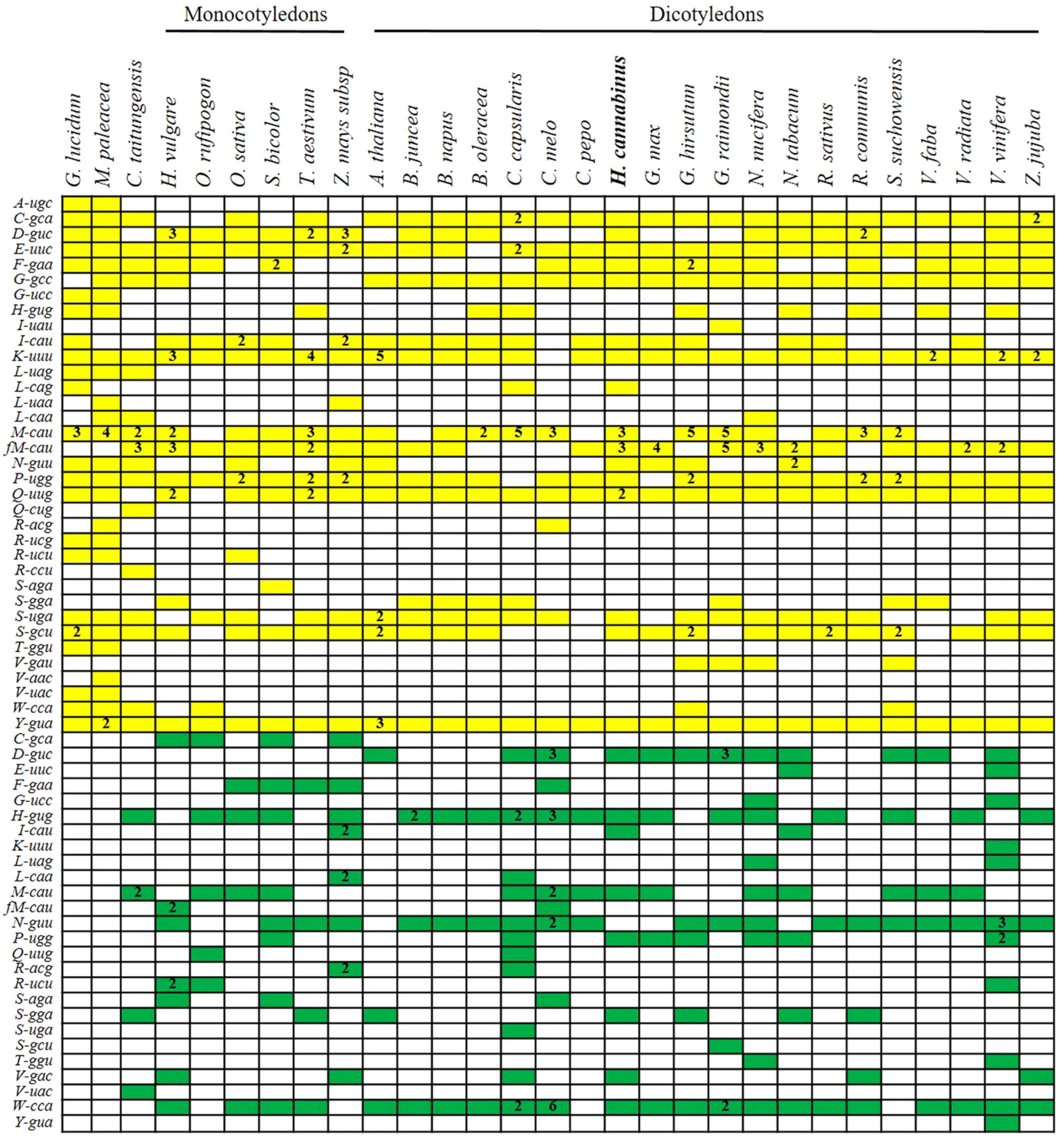


Figure 3. Analysis of conservative gene clusters between the kenaf mt genome and other higher plant mt genomes.

and gymnosperms, and bryophytes was chosen as the outgroup. The chloroplast-derived sequences and non-protein-coding sequences were removed before blasting against the other mt genomes. First, these mt functional genes were concatenated in a head-to-tail format. Maximum likelihood method was used to complete the phylogenetic tree analysis. As shown in Fig. 5, the *Hibiscus cannabinus* and *Gossypium* species belonging to the *Malvaceae* family were classified into one clade with a high bootstrap support value of 100. In addition, the species share a high sequence similarity, as supported by the higher bootstrap support values. Species belonging to different groups share less sequence similarity and have reduced bootstrap support value. The phylogenetic tree strongly supported the separation of monocot plants and dicot plants, and the separation of angiosperms



				Monocotyledons						Dicotyledons																			
	G. lucidum	*M. paleacea*	*C. taitungensis*	*H. vulgare*	*O. rufipogon*	*O. sativa*	*S. bicolor*	*T. aestivum*	*Z. mays subsp*	*A. thaliana*	*B. juncea*	*B. napus*	*B. oleracea*	*C. capsularis*	*C. melo*	*C. pepo*	***H. cannabinus***	*G. max*	*G. hirsutum*	*G. raimondii*	*N. nucifera*	*N. tabacum*	*R. sativus*	*R. communis*	*S. suchowensis*	*V. faba*	*V. radiata*	*V. vinifera*	*Z. jujuba*
A-ugc	🟨	🟨																											
C-gca	🟨	🟨	🟨			🟨		🟨		🟨	🟨	🟨	🟨	2	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	2
D-guc	🟨	🟨		3	🟨	🟨	🟨	2	3		🟨	🟨	🟨				🟨				🟨	🟨	🟨	2				🟨	🟨
E-uuc	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	2	🟨	🟨	🟨		2	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
F-gaa	🟨	🟨	🟨	🟨	🟨		2								🟨	🟨	🟨	🟨	2	🟨	🟨			🟨		🟨	🟨	🟨	🟨
G-gcc		🟨	🟨	🟨						🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
G-ucc	🟨	🟨																											
H-gug	🟨	🟨						🟨					🟨	🟨					🟨			🟨		🟨		🟨		🟨	
I-uau																				🟨									
I-cau	🟨			🟨	🟨	2	🟨		2	🟨	🟨	🟨	🟨	🟨		🟨	🟨	🟨	🟨			🟨	🟨				🟨		
K-uuu	🟨	🟨	🟨	3	🟨	🟨	🟨	4	🟨	5	🟨	🟨	🟨	🟨		🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	2	🟨	2	2
L-uag	🟨	🟨	🟨																										
L-cag	🟨													🟨			🟨												
L-uaa		🟨							🟨																				
L-caa		🟨	🟨																		🟨								
M-cau	3	4	2	2		🟨	🟨	3	🟨	🟨		🟨	2	5	3		3		5	5	🟨		🟨	3	2				
fM-cau			3	3	🟨	🟨	🟨	2	🟨	🟨	🟨	🟨				🟨	3	4		5	3	2	🟨		🟨	🟨	2	2	🟨
N-guu	🟨	🟨	🟨			🟨			🟨	🟨							🟨	🟨	🟨			2							
P-ugg	🟨	🟨	🟨	🟨	🟨	2	🟨	2	2	🟨	🟨	🟨	🟨		🟨	🟨	🟨		2	🟨	🟨	🟨	🟨	2	2	🟨	🟨	🟨	🟨
Q-uug	🟨	🟨		2		🟨	🟨	2	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	2	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
Q-cug			🟨																										
R-acg		🟨													🟨														
R-ucg	🟨	🟨																											
R-ucu	🟨	🟨				🟨																							
R-ccu			🟨																										
S-aga							🟨																						
S-gga				🟨							🟨	🟨	🟨	🟨						🟨					🟨	🟨			
S-uga	🟨	🟨	🟨		🟨	🟨		🟨	🟨	2	🟨	🟨	🟨	🟨	🟨		🟨		🟨	🟨	🟨	🟨	🟨	🟨				🟨	🟨
S-gcu	2	🟨	🟨	🟨		🟨	🟨	🟨	🟨	2	🟨	🟨	🟨				🟨	🟨	2		🟨	🟨	2	🟨	2		🟨	🟨	🟨
T-ggu	🟨	🟨																											
V-gau																			🟨	🟨	🟨				🟨				
V-aac		🟨																											
V-uac	🟨	🟨																											
W-cca	🟨	🟨	🟨		🟨														🟨						🟨				
Y-gua	🟨	2	🟨	🟨	🟨	🟨	🟨	🟨	🟨	3	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨	🟨
C-gca				🟩	🟩		🟩		🟩																				
D-guc										🟩				🟩	3		🟩	🟩	🟩	3	🟩	🟩			🟩	🟩		🟩	
E-uuc																						🟩						🟩	
F-gaa						🟩	🟩	🟩	🟩						🟩														
G-ucc																					🟩							🟩	
H-gug			🟩		🟩	🟩	🟩		🟩		2	🟩	🟩	2	3	🟩	🟩	🟩		🟩	🟩		🟩		🟩		🟩		🟩
I-cau									2								🟩					🟩							
K-uuu																												🟩	
L-uag																					🟩							🟩	
L-caa									2					🟩															
M-cau			2		🟩	🟩	🟩							🟩	2	🟩	🟩	🟩			🟩	🟩			🟩	🟩	🟩		
fM-cau				2											🟩														
N-guu				🟩			🟩	🟩	🟩		🟩	🟩	🟩	🟩	2	🟩			🟩	🟩	🟩		🟩	🟩	🟩	🟩	🟩	3	🟩
P-ugg							🟩							🟩			🟩	🟩	🟩		🟩	🟩						2	
Q-uug					🟩									🟩															
R-acg									2					🟩															
R-ucu				2	🟩																							🟩	
S-aga				🟩			🟩								🟩														
S-gga			🟩					🟩		🟩							🟩		🟩			🟩		🟩					
S-uga														🟩															
S-gcu																				🟩									
T-ggu																					🟩							🟩	
V-gac				🟩					🟩					🟩			🟩							🟩					🟩
V-uac			🟩																										
W-cca				🟩		🟩	🟩	🟩		🟩	🟩	🟩	🟩	2	6		🟩	🟩	🟩	2	🟩	🟩	🟩	🟩		🟩	🟩	🟩	🟩
Y-gua																												🟩	

Figure 4. tRNA distribution map in plant mitochondrial genomes. Yellow boxes and green boxes represent mt tRNA genes and chloroplast-like tRNA genes with one copy in plant mtDNA, respectively. The numbers in the cells represent the copy numbers in the plant mtDNA. Blank boxes indicate that the tRNA gene is absent.

from gymnosperms. Additionally, the evolutionary relationship of these 28 plant species was analysed using the plant taxonomy method and used to construct an NCBI taxonomy common tree (Fig. 6). The phylogenetic relationships based on mt genome homologous sequences are consistent with the species taxonomy despite the exceptional variability among these mt genomes.

To further explore the utility of these mt genes in phylogenetic reconstruction, twenty-two mt genes were divided into five groups according to the function of their proteins (Supplementary Table S3), and the genes in each group were assembled in a head-to-tail arrangement. Among the five groups of phylogenetic trees, the set of mitochondrial complex I genes was congruent with a previous reconstruction based on 22 functionally related genes (Supplementary Fig. S4). The trees of mitochondrial complex III and complex IV reconstructing the divergence of monocots and dicots resulted in topologies that differed from those obtained by the previous reconstruction base on 22 functionally related genes, but the species fell into clades that belonged to the same family (Supplementary Figs S5, S6). The phylogenetic trees of the mitochondrial complex V and cytochrome c biogenesis genes revealed evolution relationships that slightly differed from those obtained with the previous reconstruction base on 22 functional related genes (Supplementary Figs S7, S8). Therefore, the phylogenetic analysis base on the function of mitochondrial genes revealed slightly different topologies, but the species fell into clades that were consistent with their family designations. In contrast, the phylogenetic tree based on the whole set of genes was congruent with the species taxonomic tree (Fig. 6).

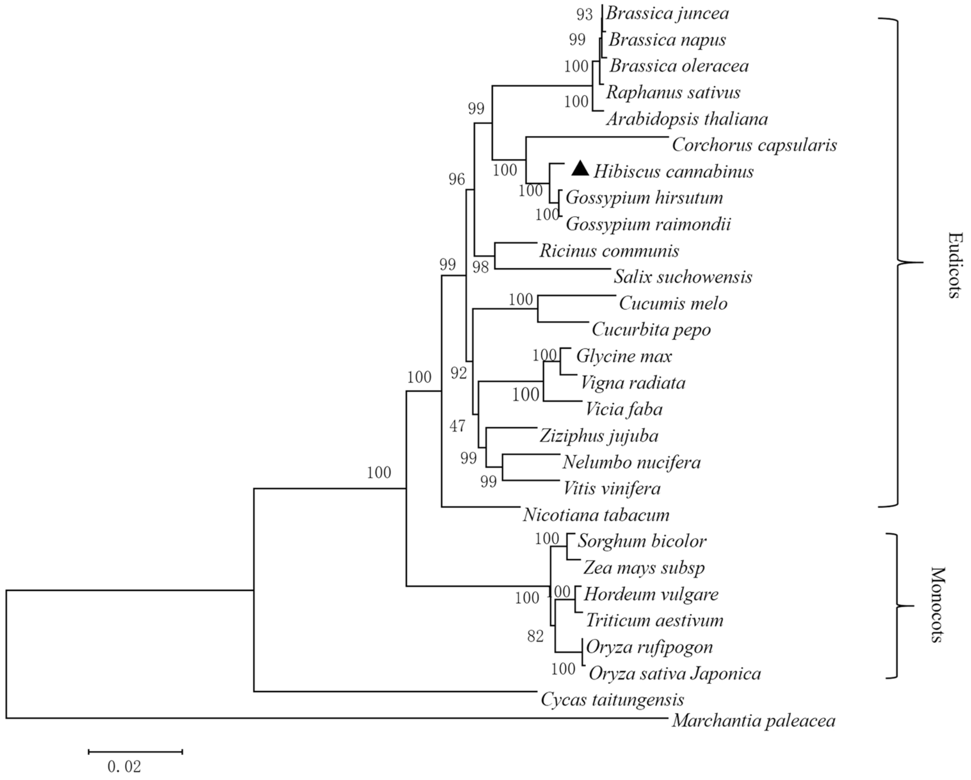


Figure 5. The original phylogenetic tree of 22 functionally related genes. The genes used are listed in Table S2 and include 17 respiratory complex genes, four cytochrome c biogenesis genes and a *cob* gene, and the tree was rooted with *Marchantia paleacea*.

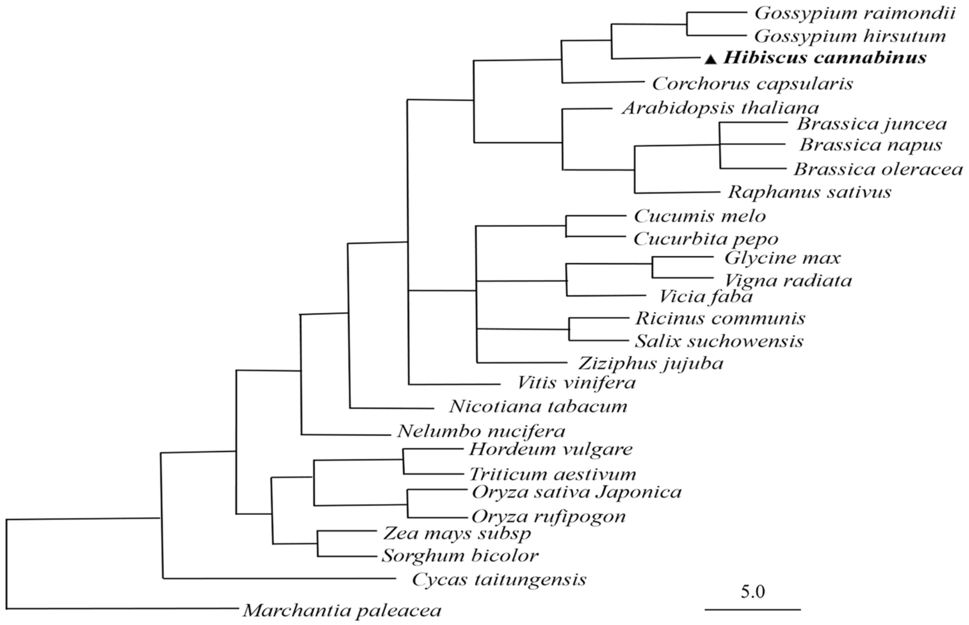


Figure 6. NCBI taxonomy common tree of 28 analysed species.

Discussion

Third-generation SMRT sequencing technology based on the PacBio RS II platform can produce substantially longer reads (>5 k/read) than second generation sequencing[28]. Which also can be used to closed genome gaps, whole-genome sequencing projects for any species[29–35] and non-SNP DNA variations identification[36]. In our study, the SMRT sequencing technology was used to sequence the whole mt genome sequence of kenaf (*Hibiscus* *cannabinus)*. We obtained the kenaf mt sequence with a high accuracy, and the genome size was 569,915 bp. The longest read was 32 kb, which is much longer than the usual reads obtained using other sequencing technologies.

Characteristics of plant mitochondrial genes. Plant mtDNAs primarily comprise of protein-coding genes, tRNAs and rRNAs. The kenaf mt genes included only 36 of the 41 protein-coding genes present in ancestral land plant mt genomes[27], indicating that several protein-coding genes were lost or transferred to other organelles during the evolution of kenaf mitochondria. The frequent loss and functional transfer of ribosomal protein genes and succinate dehydrogenase (*sdh*) genes to the nuclear genome were the main causes of the variable gene contents among the plant mt genomes. This finding is consistent with previous results that have been confirmed by Southern blot hybridization[21]. Our results revealed the presence of only *sdh4* in the kenaf mt genome, while both *sdh3* and *sdh4* were identified in the closely related *Gossypium* species[37]. Thus, the presence of the succinate dehydrogenase genes is highly variable, even among evolutionary close angiosperm species. However, two *atp9* copies were identified in the kenaf mt genome, which may have resulted from HGT events or mtDNA recombination during the evolution of the kenaf mt genome.

Repeat sequences in the genome. The mt genomes of land plants, particularly angiosperms, are frequently characterized by repeat sequences[38], which could explain most of the variation in the mt genome size. Moreover, these sequences are sites of intragenomic recombination, underlining the evolutionary changes in the mt genome organization in *vivo*[39,40]. Tandem simple and scattered repeat sequences are extensively found across plant mt genomes and exhibit high levels of polymorphisms[41,42]. The repetitive sequences in the *Cucumis melo* mt genome had a size of 2,738 kb and comprised primarily small repeats, accounting for 42.7% of the mt DNA[13]. In contrast, other genomes contain fewer larger segmental duplications[41,43]. The *Vitis* mt genome (773 kb) has only 6.8% repetitive DNA sequences[11], while the moderately sized Legume *vign*a genome (401 kb) has fewer and smaller repeats that account for 2.7% of the mt genome. These data suggest that the genome size is not a good indicator of repeat content in angiosperm mt genomes. In the present study, the repetitive structure of the kenaf mt genome accounted for 11.71% of the genome. These sequences are poorly conserved across species and have a high proportion of smaller repeats, indicating that the increased size of the kenaf mt genome was primarily due to the duplication of short sequences.

Introns in the mitochondrial genome. Land plant mt genomes contain a large and variable number of introns, ranging from 19 in *Silene latifolia*[44] to 34 in the hornwort *Phaeoceros laevis*[45]. According to the present study, the kenaf mt genome retained 25 introns, disrupting 9 protein genes, which is consistent with the common ancestor theory of gymnosperms and angiosperms[27] and suggests that the introns were lost or gained during plant evolution. However, *cis*-splicing is ubiquitous in most introns of seed plants, while the mt genes *nad1, nad2* and *nad5* evolved a split structure that requires *trans*-splicing and were highly consistent with the sequenced angiosperm mt genomes[27]. Thus, the transcription process of introns was conserved among the angiosperm mt genomes.

Conservation of gene clusters. Genome recombination can disrupt clusters, while multiple recombination events can generate similar syntenic gene clusters, leading to vast differences in the gene order among plant mt genomes[46]. In general, evolutionarily close species have more similar gene orders and clusters. Two gene clusters (*rrn18-rrn5* and *nad5-nad1-matR*) are conserved in all land plant mt genomes, and may date back to the original plant mt genomes of liverworts, mosses, and most charophytes[27]. In contrast, the *rps12-nad3* and *(rps19)-rps3-rpl16* gene clusters were evolutionary conserved in most land plant mt genomes but absent from *M. paleacea* and *S. suchowensis*, respectively[47,48]. Two other gene clusters (*rps10-cox1* and *sdh4-cox3-(atp8)*) were specifically conserved in dicots, except for *Brassica*[49]. The cluster of *atp4-nad4L* exists in all surveyed dicots, except for the species of *Gossypium*, *H.cannabinus* and *C. capsularis*. These exceptions were likely due to frequent recombination events during plant mt genome evolution. In our study, the characteristics of the gene clusters in the kenaf mt genome were consistent with the general conservation of most dicotyledon, indicating that the gene clusters in the kenaf mt genome were more conserved during plant mt genome evolution.

DNA transfer in the mitochondrial genome. HGT is thought to be the main process during the acquisition of exogenous sequences[50–52]. The transfer of DNA sequences among plastid, nuclear and mt genomes is a common phenomenon that has been observed in the fully sequenced mt genomes of land plants[53,54]. Although the amount of plastid DNA in the mt genome is 3–6% in most examined species, plastid DNA varies from 2 kb (0.5%) in the Legume *Vigna* to 113 kb (11.5%) in *Cucurbita pepo*[12]. In many cases, these plastid-to-mitochondrion transfers have resulted in the insertion of plastid genes into the mt genome, but most genes are clearly nonfunctional[53]. Occasionally, the only plastid genes that are transferred into the mt genome and remain functional encode tRNAs. The absence of chloroplast-derived tRNAs from liverworts, moss, hornworts, and bryophytes indicates that DNA transfer from the chloroplast genome to the mt genome might have occurred after the divergence of gymnosperms and angiosperms. Eight chloroplast-derived tRNAs identified in the kenaf mt genome can be traced to the retention of an earlier HGT event.

The plastid-derived *trnW* (GTT) and *trnH* (GTG) genes are frequently observed in angiosperms but are absent from *C. taitungensis*, indicating that these tRNA genes may have been transferred after the separation of angiosperms. Identifying the numbers and types of tRNA genes in the kenaf mt genome may be helpful for evaluating the origin and evolution of tRNA genes in higher plants. These results suggest that the intracellular transfer of tRNA and ribosomal genes from the chloroplast to the mitochondria was a frequent process.

Conclusion

Plant mt genomes are intriguing due to their highly conserved genic content and slow rate of genic evolution. In contrast, features, such as the genomic structure, the genome size and repeat sequences, are highly variable. In this study, we determined the complete sequence of the kenaf mt genome. The comparison of the kenaf mt genomic features with those of other plant mt genomes should provided a more comprehensive understanding of mt genome evolution in higher plants. The complete mt genome of kenaf shares many common genomic characteristics with other plant mt genomes, such as the conservation of genic content, gene clusters, certain intergenic sequences and tRNA gene origin and distribution. These observations suggest that the evolution of mt genomes is consistent with the species relationships in plant taxonomy. However, the highly dynamic genome structures (genome size and gene order) suggest that the recombination of higher plant mt genomes is independent and random among species. The sequencing of the kenaf mt genome contributes to our understanding of the characteristics of the mt genome across angiosperm evolution.

Materials and Methods

Mitochondrial DNA isolation and sequencing. Mitochondria were isolated from the kenaf maintainer line UG93B and purified from 7-day-old etiolated seedlings using differential centrifugation and discontinuous (18%, 23% and 40%) Percoll density gradient centrifugation according to the methods described by Wilson and Chourey[55]. The mitochondrial DNA isolation was performed as described by Sue[56] with modifications. The purified mitochondria were lysed with cetyltrimethylammonium bromide (CTAB) supplemented with 2% polyvinylpyrrolidone and 0.7% β-mercaptoethanol (Solarbio, Beijing) at 65 °C for 30 min. The lysis solution was extracted two to three times with chloroform/isoamyl alcohol (24:1), and absolute ethyl alcohol was used to precipitate the mtDNA. DNase-free water (50 μL) was added to resuspend the DNA pellets. The integrity, quality and concentration of the UG93B mtDNA were analysed using agarose gel electrophoresis, a NanoDrop 2000 (Thermo Scientific, USA) and a Qubit fluorometer (Thermo Scientific, USA).

DNA sequencing and genome assembly. In total, 20 μg of UG93B mtDNA were randomly sheared to fragments using a Covaris S220 (Thermo Scientific, USA). Large fragments with an average size of 20 kb were purified by magnetic bead enrichment. SMRTbell templates were obtained by ligating the hairpin adaptors to the end of a double-stranded DNA molecule and removing the failed ligation products with exonuclease. An Agilent 2100 Bioanalyzer High Sensitivity Kit (Agilent Technologies, USA) was used to assess the quality of the library. Subsequently, eight SMRT cells were sequenced using P4-C2 reagents on a PacBio RS II sequencing platform[57]. The sequencing and *de novo* assembly were performed at Nextomics Biosciences Co., Ltd, Wuhan, China. The clean reads were obtained by filtering out the sequencing adapters and low-quality sequences using SMRT Analysis 2.3.0 with the default settings. The kenaf mt genome sequence was extracted from filtered reads containing both chloroplast and mt genomes. Blat[58] was used with the default parameters against the NCBI chloroplast genome data to filter reads containing chloroplast genomes, and reads with a match greater more than 90% were moved. The kenaf mt genome sequence was assembled using a Hierarchical Genome Assembly Process (HGAP) workflow, including preassembly, error correction, Celera assembly and polishing using Quiver[59]. The long reads were selected as "seed" reads, to recruit all other subreads and construct highly accurate preassembled reads using a directed acyclic graph-based consensus procedure. This procedure was followed by assembly using off-the-shelf long-read assemblers. A basic local alignment with successive refinement (BLASR) was used to align the short reads to the seed reads and improve the sequence accuracy[60]. The Celera assembler software and overlap-layout-consensus (OLC) algorithm were used to assemble all corrected contigs[61]. Finally, Quiver was used to improve the site-specific consensus and generate the gap-free kenaf mt genome. Additionally, a specific prime pair was designed to verify the circular mitochondrial genome of kenaf (Supplementary Fig. S1).

Genome annotations and analyses. The mt genomes were annotated using BLASTn and MITOFY[12] using previous angiosperm mt genes to query sequences in the NCBI database (https://www.blast.ncbi.nlm.nih.gov). The tRNA genes were identified using the tRNA scan-SE software (http://lowelab.ucsc.edu/tRNAscan-SE/). ORFs that contained more than 100 amino-acid residues and started with methionine were predicted and annotated using ORF-Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The repeat sequences were analysed using REPuter software (http://bibiserv.techfak.uni-bielefeld.de/reputer) with the following parameters: the repeat sequence was at least 20 bp in length and the repeat identity was greater than 90%[62]. The circular map and syntenic gene cluster maps of the plant mt genomes were created using OGDRAW v1.2 (http://ogdraw.mpimp-golm.mpg.de/)[63]. The annotated genome sequence was submitted to NCBI under the GenBank accession no. MF163174.

Phylogenetic analysis. To compare the kenaf mt genome to other plant mt genomes, 28 plant mt genomes, including *Arabidopsis thaliana* (NC_001284), *Brassica juncea* (JF920288), *Brassica napus* (KP161618), *Brassica oleracea* (JF920286), *Brassica oleracea* (AP012988), *Cycas taitungensis* (NC_010303), *Citrullus lanatus* (NC_014043), *Glycine max* (JX463295), *Cucumis melo* (JF412792), *Cucurbita pepo* (NC_014050), *Gossypium raimondii* (KU317325), *Gossypium hirsutum* (JX065074), *Hordeum vulgare subsp* (AP017301), *Marchantia paleacea* (NC_001660), *Nelumbo nucifera* (KR610474), *Nicotiana tabacum* (BA000042), *Oryza rufipogon* (AP011076), *Oryza sativa japonica* (BA000029), *Raphanus sativus (JQ083668), Ricinus communis* (HQ874649), *Salix suchowensis* (NC_029317*), Sorghum bicolor* (DQ 984518), *Triticum aestivum* (AP008982), *Vicia faba* (KC189947), *Vigna radiate* (NC_015121), *Vitis vinifera* (NC_012119), *Zea mays subsp. mays* (NC_007982), and *Ziziphus jujuba* (KU187967), were downloaded from the NCBI Organelle Genome Resources database (http://www.ncbi.nlm.

nih.gov/genome/organelle/). These mt genome sequences were selected because they are available for analysis in NCBI and are clearly taxonomically classified. Phylogenetic analyses were performed using concatenated exon sequences from 22 conserved protein-coding genes (*atp1, atp4, atp6, atp8, atp9, ccmB, ccmC, ccmFc, ccmFn, cob, cox1, cox2, cox3, nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7 and nad9*) extracted from these 28 plant mt genomes. These nucleotides were aligned using ClustalW and manually modified to eliminate gaps and missing data. Finally, the maximum likelihood (ML) method was used to construct original phylogenetic trees by MEGA 6.0[64]. The bootstrap replications were performed with 1000 according to Felsenstein[65]. The evolutionary distances were computed using the Kimura 2-parameter method[66] and the tree was rooted with *Marchantia paleacea*. The NCBI taxonomy common tree was described by Federhen[67] and constructed using the online NCBI taxonomy database (https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi).

References

1. Lang, B. F., Gray, M. W. & Burger, G. Mitochondrial genome evolution and the origin of eukaryotes. *Annual Review of Genetics* **33**, 351–397 (1999).
2. Bergthorsson, U., Richardson, A. O., Young, G. J., Goertzen, L. R. & Palmer, J. D. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proc Natl Acad Sci USA* **101**, 17747–17752 (2004).
3. Alverson, A. J., Zhuo, S., Rice, D. W., Sloan, D. B. & Palmer, J. D. The mitochondrial genome of the Legume *vigna radiata* and the analysis of recombination across short mitochondrial repeats. *Plos One* **6**, e16404 (2011).
4. Unseld, M., Marienfeld, J. R., Brandt, P. & Brennicke, A. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat Genet* **15**, 57–61 (1997).
5. Kubo, T. *et al*. 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 (2000).
6. Notsu, Y. *et al*. 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 (2002).
7. 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* **31**, 5907–5916 (2003).
8. Clifton, S. W. *et al*. Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol* **136**, 3486–3503 (2004).
9. Ogihara, Y. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Res* **33**, 6235–6250 (2005).
10. Sugiyama, Y. *et al*. 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 (2005).
11. Goremykin, V. V., Salamini, F., Velasco, R. & Viola, R. Mitochondrial DNA of *Vitis Vinifera* and the issue of rampant horizontal gene transfer. *Mol Biol Evol* **26**, 99–110 (2008).
12. Alverson, A. J. *et al*. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus Lanatus* and *Cucurbita Pepo* (*Cucurbitaceae*). *Mol Biol Evol* **27**, 1436–1448 (2010).
13. Rodriguez-Moreno, L. *et al*. Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. *BMC Genomics* **12**, 424 (2011).
14. Liu, G. *et al*. The Complete mitochondrial genome of *Gossypium hirsutum* and evolutionary analysis of higher plant mitochondrial genomes. *Plos One* **8**, e69476 (2013).
15. Li, S. *et al*. Construction and initial analysis of five fosmid libraries of mitochondrial genomes of cotton (*Gossypium*). *Chinese. Sci Bull* **36**, 4608–4616 (2013).
16. Chang, S. *et al*. The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels. *Plos One* **8**, e56502 (2013).
17. Chang, S. *et al*. The mitochondrial genome of *raphanus sativus* and gene evolution of cruciferous mitochondrial types. *J Genet Genomics* **40**, 117–126 (2013).
18. Negruk, V. Mitochondrial genome sequence of the Legume *Vicia faba*. *Front Plant Sci.* **4**, 128 (2013).
19. Tanaka, Y., Tsuda, M., Yasumoto, K., Terachi, T. & Yamagishi, H. The complete mitochondrial genome sequence of *Brassica oleracea* and analysis of coexisting mitotypes. *Curr Genet* **60**, 277–284 (2014).
20. Sloan, D. B. *et al*. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *Plos Biol* **10**, e1001241 (2012).
21. Adams, K. L., Qiu, Y. L., Stoutemyer, M. & Palmer, J. D. 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 USA* **99**, 9905–9912 (2002).
22. Hanson, M. R. Interactions of mitochondrial and nuclear genes that affect male gametophyte development. *The Plant Cell Online* **16**, S154–S169 (2004).
23. Chen, L. & Liu, Y. G. Male sterility and fertility restoration in crops. *Annu Rev Plant Biol.* **65**, 579–606 (2014).
24. Gualberto, J. M. & Newton, K. J. Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu Rev Plant Biol.* **68**, 225–252 (2017).
25. Vaughn, J. C., Mason, M. T., Sper-Whitis, G. L., Kuhlman, P. & Palmer, J. D. Fungal Origin by Horizontal transfer of a plant mitochondrial group I intron in the chimeric *coxI* gene of *Peperomia*. *J Mol Evol.* **41**, 563–572 (1995).
26. Monti, A. & Alexopoulou, E. Kenaf: A multi-purpose crop for several industrial applications (ed. Monti, A.) 105–143 (Springer-Verlag, 2013).
27. Mower, J. P., Sloan, D. B. & Alverson, A. J. Plant mitochondrial genome diversity: the genomics revolution (ed. Mower, P.) 123–144 (Springer Vienna, 2012).
28. Quail, M. A. *et al*. A tale of three next generation sequencing platforms: comparison of ion torrent, pacific biosciences and illumina miseq sequencers. *BMC Genomics* **13**, 341 (2012).
29. Gui, S. *et al*. The Mitochondrial genome map of *Nelumbo nucifera* reveals ancient evolutionary features. *Sci Rep* **6**, 30158 (2016).
30. Rothfels, C. J., Pryer, K. M. & Li, F. W. Next-generation polyploid phylogenetics: rapid resolution of hybrid polyploid complexes using pacbio single-molecule sequencing. *New Phytol* **213**, 413–429 (2017).
31. Treffer, R. & Deckert, V. Recent advances in single-molecule sequencing. *Curr Opin Biotechnol* **21**, 4–11 (2010).
32. Eid, J. *et al*. Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).
33. Pushkarev, D., Neff, N. F. & Quake, S. R. Single-molecule sequencing of an individual human genome. *Nat Biotechnol* **27**, 847–850 (2009).
34. Liao, Y., Lin, S. & Lin, H. Completing bacterial genome assemblies: strategy and performance comparisons. *Sci Rep* **5**, 8747 (2015).
35. Korlach, J. *et al*. Real-time DNA sequencing from single polymerase molecules. *Methods Enzymol* **472**, 431–455 (2010).
36. Ritz, A. *et al*. Characterization of structural variants with single molecule and hybrid sequencing approaches. *Bioinformatics.* **30**, 3458–3466 (2014).

37. Tang, M. *et al*. Rapid evolutionary divergence of *Gossypium barbadense* and G.*hirsutum* mitochondrial genomes. *BMC Genomics* **16**, 770 (2015).
38. Kempken, F., Knoop, V., Volkmar, U., Hecht, J., & Grewe, F. Plant mitochondria (ed. Kempken, F.) 3–29 (Springer, 2011).
39. Abdelnoor, R. V. *et al*. Mitochondrial genome dynamics in plants and animals: convergent gene fusions of a *muts* homologue. *J Mol Evol* **63**, 165–173 (2006).
40. Gualberto, J. M. *et al*. The plant mitochondrial genome: dynamics and maintenance. *Biochimie* **100**, 107–120 (2014).
41. Iorizzo, M. *et al*. De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. *BMC Plant Biol* **12**, 61 (2012).
42. Fujii, S., Kazama, T., Yamada, M. & Toriyama, K. Discovery of global genomic re-organization based on comparison of two newly sequenced rice mitochondrial genomes with cytoplasmic male sterility-related genes. *BMC Genomics* **11**, 209 (2010).
43. Allen, J. O. *et al*. Comparisons among two fertile and three male-sterile mitochondrial genomes of *maiz*e. *Genetics.* **177**, 1173–1192 (2007).
44. Sloan, D. B. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. *BMC Evol Biol* **10**, 274 (2010).
45. Xue, J. Y., Liu, Y., Li, L., Wang, B. & Qiu, Y. L. The complete mitochondrial genome sequence of the *hornwort Phaeoceros* laevis: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in *hornworts*. *Curr Genet* **56**, 53–61 (2010).
46. Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. & Maltsev, N. The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci USA* **96**, 2896–2901 (1999).
47. Bonavita, S. & Regina, T. M. The evolutionary conservation of*rps3* introns and *rps19-rps3-rpl16* gene cluster in *Adiantum capillus-veneri*s mitochondria. *Curr Genet* **62**, 173–184 (2016).
48. Dias, S. M., Siqueira, S. F., Lejeune, B. & de Souza, A. P. Identification and characterization of the trnS/pseudo-trnA/*nad3/rps12* gene cluster from *Coix lacryma-jobi* L: organization, transcription and RNA editing. *Plant Sci* **158**, 97–105 (2000).
49. Siqueira, S. F. *et al*. Transcription of succinate dehydrogenase subunit 4 (*sdh*4) gene in potato: detection of extensive RNA Editing and co-transcription with cytochrome oxidase subunit III *(cox*3) gene. *Curr Genet* **41**, 282–289 (2002).
50. Warren, J. M., Simmons, M. P., Wu, Z. & Sloan, D. B. Linear plasmids and the rate of sequence evolution in plant mitochondrial genomes. *Genome Biol Evol* **8**, 364–374 (2016).
51. Wu, B., Buljic, A. & Hao, W. Extensive horizontal transfer and homologous recombination generate highly chimeric mitochondrial genomes in yeast. *Mol Biol Evol.* **32**, 2559–2570 (2015).
52. Rice, D. W. *et al*. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* **342**, 1468–1473 (2013).
53. Straub, S. C., Cronn, R. C., Edwards, C., Fishbein, M. & Liston, A. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (*Apocynaceae*). *Genome Biol Evol* **5**, 1872–1885 (2013).
54. Bergthorsson, U., Adams, K. L., Thomason, B. & Palmer, J. D. Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* **424**, 197–201 (2003).
55. Wilson, A. J. & Chourey, P. S. A Rapid inexpensive method for the isolation of restrictable mitochondrial DNA from various plant sources. *Plant Cell Rep* **3**, 237–239 (1984).
56. Sue Porebski, L. G. B. A. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter* **15**, 8–15 (1997).
57. Rhoads, A. & Au, K. F. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* **13**, 278–289 (2015).
58. Kent, W. J. BLAT-the BLAST-like alignment tool. *Genome Res* **12**, 656–664 (2002).
59. Chin, C. S. *et al*. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* **10**, 563–569 (2013).
60. Chaisson, M. J. & Tesler, G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* **13**, 238 (2012).
61. Myers, E. W. *et al*. A Whole-genome assembly of drosophila. *Science* **287**, 2196–2204 (2000).
62. Kurtz, S. *et al*. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* **29**, 4633–4642 (2001).
63. Lohse, M., Drechsel, O., Kahlau, S. & Bock, R. Organellar GenomeDRAW-a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* **41**, 575–581 (2013).
64. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729 (2013).
65. Felsenstein J, Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783–791 (1985).
66. Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120 (1980).
67. S. Federhen, The NCBI Taxonomy database. *Nucleic Acids Research* **40**, 136–143 (2011)

Acknowledgements

This study was supported by the National Science Foundation of China (No. 31571719, No. 31660430) and the Chinese Postdoctoral Science Foundation (No. 2016M592608). The authors thank the Wuhan NextOmics Biotechnology Limited Company (China) for their help in sequencing the kenaf mtDNA.

Author Contributions

H.W. and R.Z. initiated the experiment. X.L. conducted the experiment and drafted the manuscript. Y.Z. isolated the kenaf mtDNA. X.K., B.Z. and D.L. assisted with the experiment. A.K. and M.H.K. revised the manuscript and P.C. provided suggestions and edits for the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-30297-w.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.