



Article

The Complete Chloroplast Genome Sequence of Tree of Heaven (*Ailanthus altissima* (Mill.) (Sapindales: Simaroubaceae), an Important Pantropical Tree

Joshat K. Saina^{1,2,3,4}, Zhi-Zhong Li^{2,3}, Andrew W. Gichira^{2,3,4} and Yi-Ying Liao^{1,*}

¹ Fairy Lake Botanical Garden, Shenzhen & Chinese Academy of Sciences, Shenzhen 518004, China; jksaina@wbpcas.cn

² Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China; wbg_georgelee@163.com (Z.-Z.L.); gichira@wbpcas.cn (A.W.G.)

³ University of Chinese Academy of Sciences, Beijing 100049, China

⁴ Sino-African Joint Research Center, Chinese Academy of Sciences, Wuhan 430074, China

* Correspondence: liaoyiying666@163.com; Tel.: +86-150-1949-8243

Received: 31 January 2018; Accepted: 16 March 2018; Published: 21 March 2018



Abstract: *Ailanthus altissima* (Mill.) Swingle (Simaroubaceae) is a deciduous tree widely distributed throughout temperate regions in China, hence suitable for genetic diversity and evolutionary studies. Previous studies in *A. altissima* have mainly focused on its biological activities, genetic diversity and genetic structure. However, until now there is no published report regarding genome of this plant species or Simaroubaceae family. Therefore, in this paper, we first characterized *A. altissima* complete chloroplast genome sequence. The tree of heaven chloroplast genome was found to be a circular molecule 160,815 base pairs (bp) in size and possess a quadripartite structure. The *A. altissima* chloroplast genome contains 113 unique genes of which 79 and 30 are protein coding and transfer RNA (tRNA) genes respectively and also 4 ribosomal RNA genes (rRNA) with overall GC content of 37.6%. Microsatellite marker detection identified A/T mononucleotides as majority SSRs in all the seven analyzed genomes. Repeat analyses of seven Sapindales revealed a total of 49 repeats in *A. altissima*, *Rhus chinensis*, *Dodonaea viscosa*, *Leitneria floridana*, while *Azadirachta indica*, *Boswellia sacra*, and *Citrus aurantiifolia* had a total of 48 repeats. The phylogenetic analysis using protein coding genes revealed that *A. altissima* is a sister to *Leitneria floridana* and also suggested that Simaroubaceae is a sister to Rutaceae family. The genome information reported here could be further applied for evolution and invasion, population genetics, and molecular studies in this plant species and family.

Keywords: *Ailanthus altissima*; chloroplast genome; microsatellites; Simaroubaceae; Sapindales

1. Introduction

Ailanthus altissima (Mill.) Swingle, a deciduous tree in the Simaroubaceae family, is widely distributed throughout temperate regions in China. It grows rapidly reaching heights of 15 m (49ft) in 25 years and can tolerate various levels of extreme environments (e.g., low temperatures, sterile soils, arid land). Besides, it reproduces through sexual (seeds disperse by wind) or asexual (sprouts) methods [1]. Two hundred years ago it was brought to Europe and North America. *A. altissima* being an early colonizer can survive high levels of natural or human disturbance [2]. Therefore, in recent years, it is commonly known as an exotic invasive tree developed into an invasive species expanding on all continents except Antarctica [1]. While previous studies in *A. altissima* have mainly focused on discovering the biological features of this plant to prevent its expansion, there is limited information to

understand the impact of genetic diversity and evolution. Thus, genomic information of *A. altissima* is essential for further molecular studies, identification, and evolutionary studies.

Many studies have analyzed the genetic diversity of *A. altissima* using various markers, for example, chloroplast DNA [2,3], microsatellite primer [4,5]. These studies provided a detailed series of information about genetic structure and genetic diversity of *A. altissima* in native and invasive area. However, to understand the genetic diversity and population structure within *A. altissima* natural populations, more genetic resources are required.

It is well known that chloroplasts (cp) are key organelles in plants, with crucial functions in the photosynthesis and biosynthesis [6]. Current research shows that chloroplast genomes in angiosperms have highly conserved structure, gene content, organization, compared with either nuclear or mitochondrial genomes [7,8]. In general, cp genomes in angiosperms have circular structure consisting of two inverted repeat regions (IRa and IRb) that divides a large–single-copy (LSC) and a small–single-copy (SSC) regions [9]. Nevertheless, mutations, duplications, arrangements and gene loss have been observed, including the loss of the inverted repeat region in leguminous plants [7,10–12]. Some studies have applied plant plastomes to study population genetic analyses and basal phylogenetic relationships at different taxonomic levels [13], also to investigate the functional and structural evolution in plants [14–16]. At present, more cp genomes have been sequenced as a result of next-generation sequencing technologies advancement resulting in low sequencing costs.

More than 800 sequenced plastomes from various land plants have boosted our understanding of intracellular gene transfer, conservation, diversity, and genetic basis [17]. Although cp genomes have been sequenced in many trees such as *Castanea mollissima* [18]), *Liriodendron tulipifera* [19], *Eucalyptus globules* [20], and *Larix deciduas* [21], the plastome of *Leitneria floridana* (GenBank NC_030482) a member of Simaroubaceae has been sequenced but no analysis has been published at present despite the family containing many high economic value trees. Regardless of its potential use in crop or tree species improvement, studies on invasive species such as *A. altissima* which is also an important economic tree in the North China are too few. Chloroplast genome sequencing in invasive species could bring insights into evolutionary aspects in stress-tolerance related trait and genetic variation.

Simple sequence repeat (SSR) also called microsatellite markers are known to be more informative and versatile DNA-based markers used in plant genetic research [22]. These DNA markers are reliable molecular tools that can be used to examine plants genetic variation. SSR loci are evenly distributed and very abundant in angiosperms plastomes [23,24]. Chloroplast microsatellites are typically mononucleotide tandem repeats, and SSR in the fully sequenced genome could be used in plant species identification and diversity analysis. CpSSR in the fully sequenced plants plastomes such as; orchid genus *Chiloglottis* [25], *Cryptomeria japonica* [26], *Podocarpus lambertii* [27], *Actinidia chinensis* [28], have proven to be useful genetic tools in determining gene flow and population genetics of cp genomes. However, the lack of published plastome of *A. altissima* has limited the development of suitable SSR markers.

Here, we sequenced the complete chloroplast genome of *A. altissima*, and characterized its organization and structure. Furthermore, phylogenetic relationship using protein coding genes from selected species, consisting of 31 species from five families was uncovered for the Simaroubaceae family within the order Sapindales. Lastly, this resource will be used to develop SSR markers for analyzing genetic diversity and structure of several wild populations of *A. altissima*.

2. Results and Discussion

2.1. *Ailanthus altissima* Genome Size and Features

The *A. altissima* chloroplast genome has a quadripartite organizational structure with overall size of 160,815 base pairs (bp) including two copies of Inverted repeats (IRa and IRb) (27,051 bp each) separating the Large Single Copy (LSC) (88,382 bp) and Small Single Copy Region (SSC) (18,332 bp) (Figure 1). Notably, the genome content; gene order, orientation and organization of *A. altissima* is

similar to the reference genome and other sequenced Sapindales plastomes [29,30] with genome size of about 160 kb. The overall guanine-cytosine (GC) content of the whole genome is 37.6%, while the average adenine-thymine A + T content is 62.36%. The relatively higher IR GC content and A + T bias in this chloroplast have been previously reported in genomes of relative species in order Sapindales [31]. The GC content of the LSC, SSC and IR regions are 35.7, 32.2 and 42.6% respectively. Moreover, the protein coding sequences had 38.3% GC content.

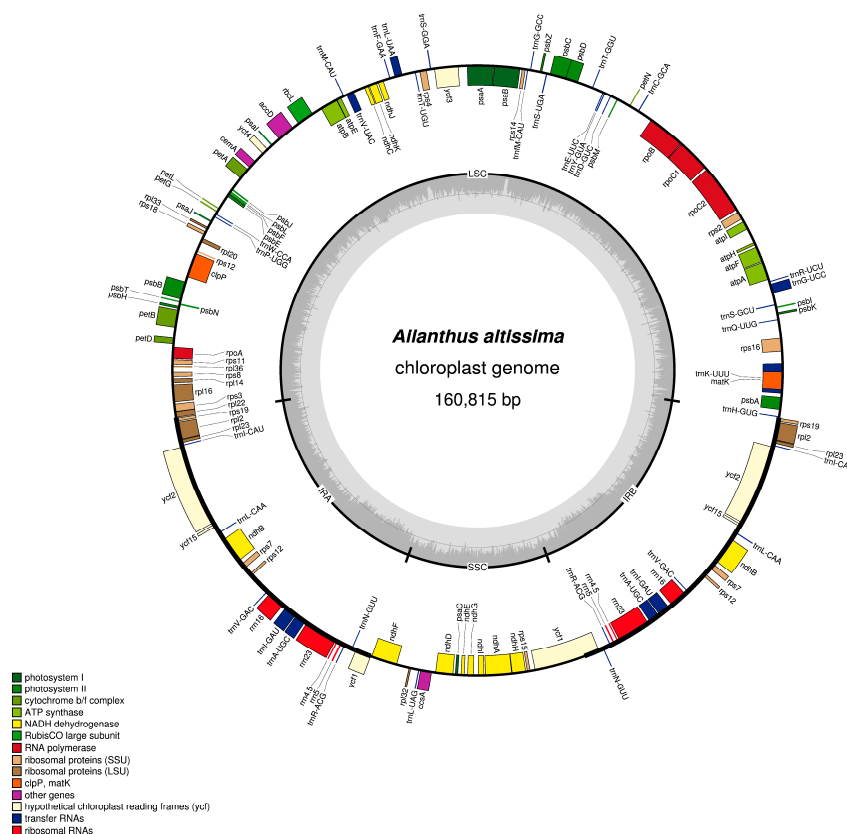


Figure 1. Circular gene map of *A. altissima* complete chloroplast genome. Genes drawn on the outside of the circle are transcribed clockwise, whereas those inside are transcribed counter-clockwise. The light gray in the inner circle corresponds to AT content, while the darker gray corresponds to the GC content. Large single copy (LSC), Inverted repeats (IRa and IRb), and Small single copy (SSC) are indicated.

The tree of heaven (*A. altissima*) chloroplast genome harbored a total of 113 different genes, comprising 79 protein coding genes (PCGs), 30 transfer RNA (tRNA) genes, and four ribosomal RNA (rRNA) genes (Table 1). All the 77 PCGs started with the common initiation codon (ATG), but *rps19* and *ndhD* genes started with alternative codons GTG and ACG respectively, this unusual initiator codons have been observed to be common in other angiosperm chloroplast genomes [32–34]. Of the 79 protein coding sequences, 60 are located in the LSC, 11 in the SSC and eight genes were duplicated in the IR region, while 22 tRNA genes were found in LSC, seven replicated in the IR region and one located in the SSC region.

Similar to some closely related plant species in order Sapindales, the chloroplast genome of *A. altissima* has maintained intron content [35]. Among the 113 unique genes, the *rps16*, *rpoC1*, *petB*, *rpl2*, *ycf15*, *ndhB*, *ndhA*, *atpF*, six tRNA genes (*trnG-GCC**, *trnA-UGC**, *trnL-UAA**, *trnI-GAU**, *trnK-UUU**, *trnV-UAC**) possess one intron, and *ycf3* and *clpP* genes harbored two introns. The *rps12* trans-splicing gene has two 3' end exons repeated in the IRs and the 5' end exon located in the LSC region, which is similar to that in *C. platymamma* [30], *C. aurantiifolia* [29], *Dipteronia* species [35]. The *ycf1* gene crossed the IR/SSC junction forming a pseudogene *ycf1* on the corresponding IR region. The *rps19* gene in

A. altissima was completely duplicated in the inverted repeat (IR) region, which most other chloroplast genomes have presented [29,30].

Table 1. List of genes found in *Ailanthus altissima* Chloroplast genome.

Functional Category	Group of Genes	Gene Name	Number
Self-replication	rRNA genes	<i>rrn16</i> (×2), <i>rrn23</i> (×2), <i>rrn4.5</i> (×2), <i>rrn5</i> (×2),	4
	tRNA genes	<i>trnA</i> -UGC*(×2), <i>trnC</i> -GCA, <i>trnD</i> -GUC, <i>trnE</i> -UUC, <i>trnF</i> -GAA, <i>trnG</i> -UCC, <i>trnH</i> -GUG, <i>trnI</i> -CAU(×2), <i>trnI</i> -GAU*(×2), <i>trnK</i> -UUU*, <i>trnL</i> -CAA(×2), <i>trnL</i> -UAA*, <i>trnL</i> -UAG, <i>trnG</i> -GCC*, <i>trnM</i> -CAU, <i>trnN</i> -GUU(×2), <i>trnP</i> -GGG, <i>trnP</i> -UGG, <i>trnQ</i> -UUG, <i>trnR</i> -ACG(×2), <i>trnR</i> -UCU, <i>trnS</i> -GCU, <i>trnS</i> -GGA, <i>trnS</i> -UGA, <i>trnT</i> -GGU, <i>trnT</i> -UGU, <i>trnV</i> -GAC(×2), <i>trnV</i> -UAC*, <i>trnW</i> -CCA, <i>trnY</i> -GUA	30
	Ribosomal small subunit	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i> , <i>rps11</i> , <i>rps12</i> , <i>rps14</i> , <i>rps15</i> , <i>rps16</i> *, <i>rps18</i> , <i>rps19</i>	12
	Ribosomal large subunit	<i>rpl2</i> *(×2), <i>rpl14</i> , <i>rpl16</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (×2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>	9
	DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> *, <i>rpoC2</i>	4
Photosynthesis	Large subunit of rubisco	<i>rbcl</i>	1
	Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i> , <i>ycf3</i> **	6
	Photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>	15
	NADH dehydrogenase	<i>ndhA</i> *, <i>ndhB</i> *(×2), <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>	11
	Cytochrome b/f complex	<i>petA</i> , <i>petB</i> *, <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>	6
	ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> *, <i>atpH</i> , <i>atpI</i>	6
	Other	Maturase	<i>matK</i>
	Subunit of acetyl-CoA carboxylase	<i>accD</i>	1
	Envelope membrane protein	<i>cemA</i>	1
	Protease	<i>clpP</i> **	1
	c-type cytochrome synthesis	<i>ccsA</i>	1
Functions unknown	Conserved open reading frames (<i>ycf</i>)	<i>ycf1</i> , <i>ycf2</i> (×2), <i>ycf4</i> , <i>ycf15</i> (×2)	4
Total			113

Note: * Gene with one intron, ** Genes with two introns. (×2) Genes with two copies.

2.2. IR Expansion and Contraction and Genome Rearrangement

The angiosperms chloroplast genomes are highly conserved, but slightly vary as a result of either expansion or contraction of the single-copy (SC) and IR boundary regions [36]. The expansion and contraction of the IR causes size variations and rearrangements in the SSC/IRa/IRb/LSC junctions [37]. Therefore, in this study, exact IR boundary positions and their adjacent genes of seven representative species from different families in order Sapindales were compared (Figure 2). The functional *ycf1* gene crossed the IRa/SSC boundary creating *ycf1* pseudogene fragment at the IRb region in all the genomes. Besides, *ycf1* pseudogene overlapped with the *ndhF* gene in the SSC and IRa junctions in four genomes with a stretch of 9 to 85 bp, but *ndhF* gene is located in SSC region in *L. floridana*, *R. chinensis* and *A. altissima*.

The *rpl22* gene crossed the LSC/IRb junction in all the chloroplast genomes except in *R. chinensis*. Furthermore, this gene was partially duplicated forming a pseudogene fragment at the corresponding IRA/LSC junction in *L. floridana* and *B. sacra*, but completely duplicated in *D. viscosa*. In all the seven chloroplast genomes, the *trnH-GUG* gene was located in the LSC regions, however this gene overlapped with *rpl22* gene in *D. viscosa*. The results reported here are congruent with the recent studies which showed that the *trnH-GUG* gene was situated in the LSC region in some species from order Sapindales, while the SSC/IRa border extends into the protein coding gene *ycf1* with subsequent

formation of a *ycf1* pseudogene [29,30]. Despite the seven chloroplast genomes of Sapindales having well-conserved genomic structure in terms of gene order and number, length variation of the whole chloroplast genome sequences and LSC, SSC and IR regions was detected among these genomes. This sequence variation might have been as a result of boundary expansion and contraction between the single copy and IR boundary regions among plant lineages as suggested by Wang and Messing 2011 [38].

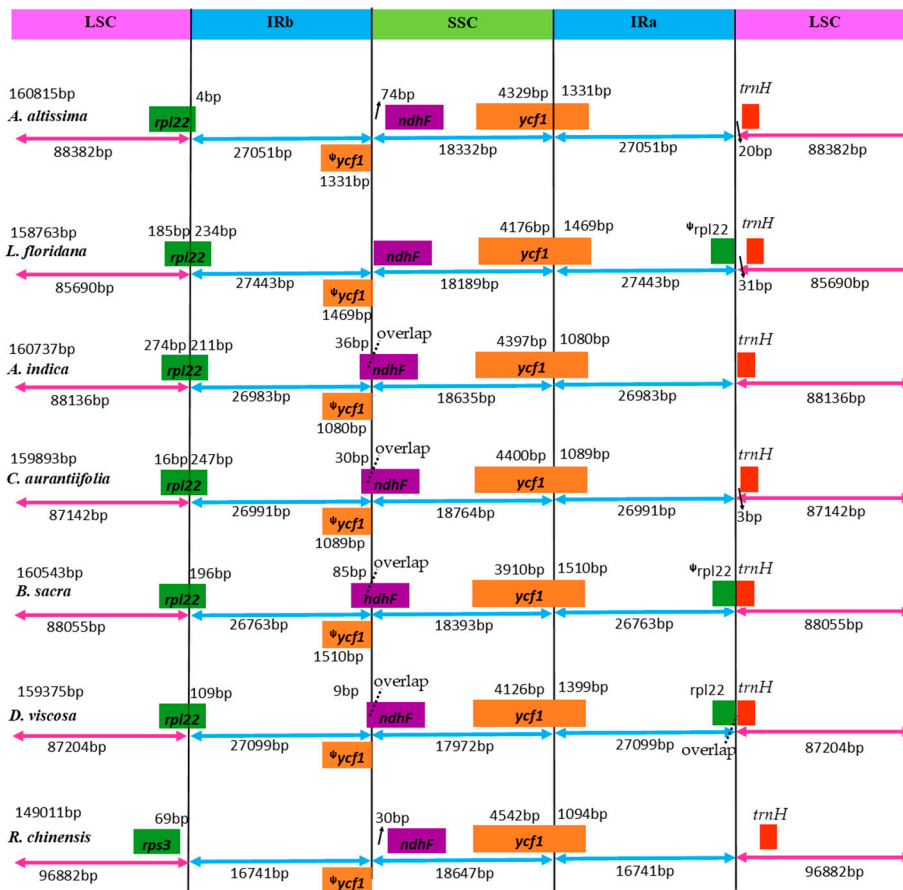


Figure 2. Comparison of IR, LSC and SSC junction positions among seven Chloroplast genomes. The features drawn are not to scale. The symbol ϕ means pseudogene created by IRb/SSC border extension into *ycf1* genes. Colored boxes for genes represent the gene position.

The mauve alignment for seven species revealed that all the genomes formed collinear blocks (LCBs). In particular, all the seven species; *A. altissima*, *Leitneria floridana*, *Azadirachta indica*, *Citrus aurantiifolia*, *Boswellia Sacra*, *Spondias bahiensis* and *Dodonaea viscosa* reveal a syntenic structure, however block two was inverted (from *rpl20* to *rbcL* genes) compared to the reference genome (*Aquilaria sinensis*). The collinear blocks of the genes including ribosomal RNA, tRNA, and protein coding genes revealed that all the seven genomes were relatively conserved with no gene rearrangement (Figure 3). Some other studies have revealed homology in genome organization and no gene rearrangements, thus our findings support their conclusions [31,39,40].

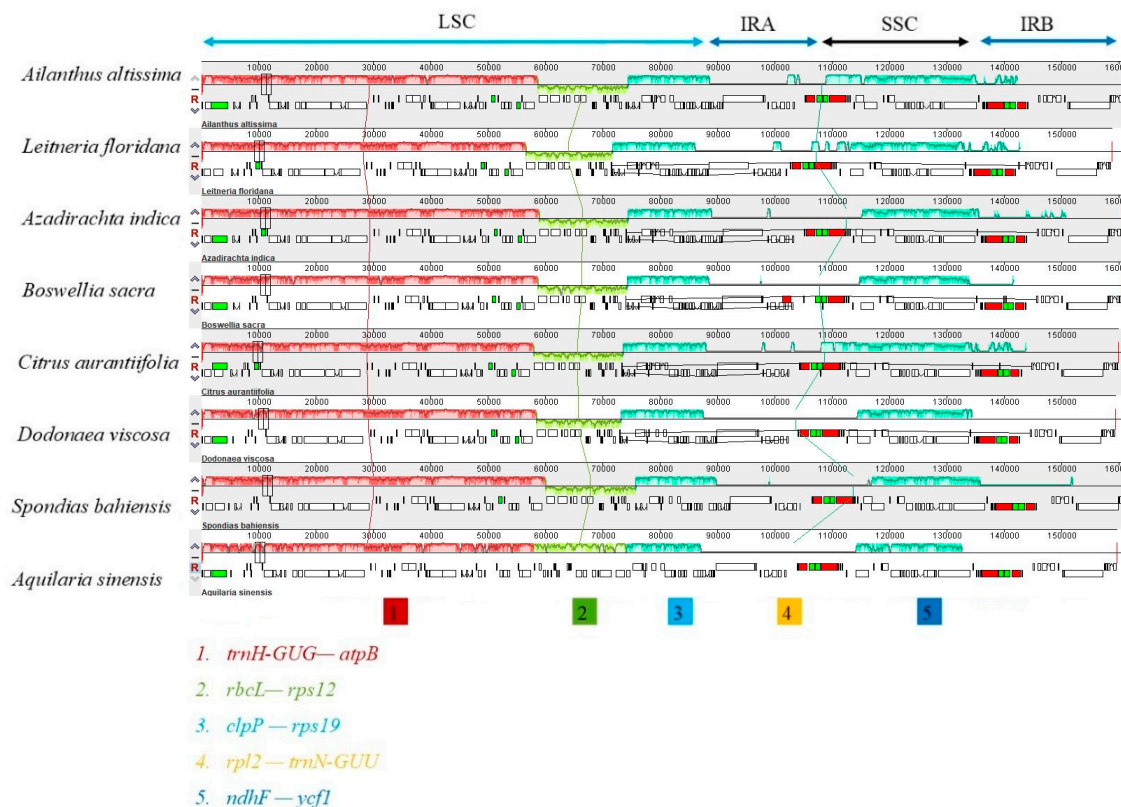


Figure 3. Gene arrangement map of seven chloroplast genomes representing families from Sapindales, and one reference species (*Aquilaria sinensis*) aligned using Mauve software. Local collinear blocks within each alignment are represented in as blocks of similar color connected with lines. Annotations of rRNA, protein coding and tRNA genes are shown in red, white and green boxes respectively.

2.3. Codon Usage and Putative RNA Editing Sites in Chloroplast Genes of *A. altissima*

In this study, we analyzed codon usage frequency and the relative synonymous codon usage (RSCU) in the *A. altissima* plastome. All the protein coding genes presented a total of 68,952 bp and 22,964 codons in *A. altissima* chloroplast genome. Of 22,964 codons, leucine (Leu) being the most abundant amino acid had a frequency of 10.56%, then isoleucine (Ile) with 8.54%, while cysteine (Cys) was rare with a proportion of 1.12% (Tables S1 and S2, Figure 4). Our study species genome is like other previously reported genomes which showed that leucine and isoleucine are more common [41–45]. Furthermore, comparable to other angiosperm chloroplast genomes, our results followed the trend of codon preference towards A/T ending which was observed in plastomes of two *Aristolochia* species [46], *Scutellaria baicalensis* [47], *Decaisnea insignis* [34], *Papaver rhoeas* and *Papaver orientale* [48] *Cinnamomum camphora* [49], and *Forsythia suspensa* [41]. All the twenty-eight A/U—ending codons had RSCU values of more than one ($RSCU > 1$), whereas the C/G—ending codons had RSCU values of less than one. Two amino acids, Methionine (Met) and tryptophan (Trp) showed no codon bias. The results for number of codons (Nc) of each protein coding gene ranged from 38.94 (*rps14* gene) to 58.37 (*clpP* gene).

The potential RNA editing sites in tree of heaven chloroplast genome was done using PREP program which revealed that most conversions at the codon positions change from serine (S) to leucine (L) (Table 2). In addition, 15 (27.78%), 39 (72.22%), and 0 editing locations were used in the first, second and third codons respectively. One RNA editing site converted the amino acid from apolar to polar (proline (P) to serine (S)). Overall, the PREP program identified a total of 54 editing sites in 21 protein coding genes, with *ndhB* and *ndhD* genes predicted to have the highest number of editing sites (9). Followed by *ndhA*, *matK*, *rpoC2*, and *rpoB* with four editing sites, whereas *ndhF* had three sites. Interestingly, fifty three of fifty four RNA editing conversions in the *A. altissima* chloroplast

genome resulted into hydrophobic products comprising; isoleucine, leucine, tryptophan, tyrosine valine, methionine, and phenylalanine. In general our results are congruent with the preceding reports which also found that most RNA editing sites led to amino acid change from polar to apolar, resulting in increase in protein hydrophobicity [41,46,50].

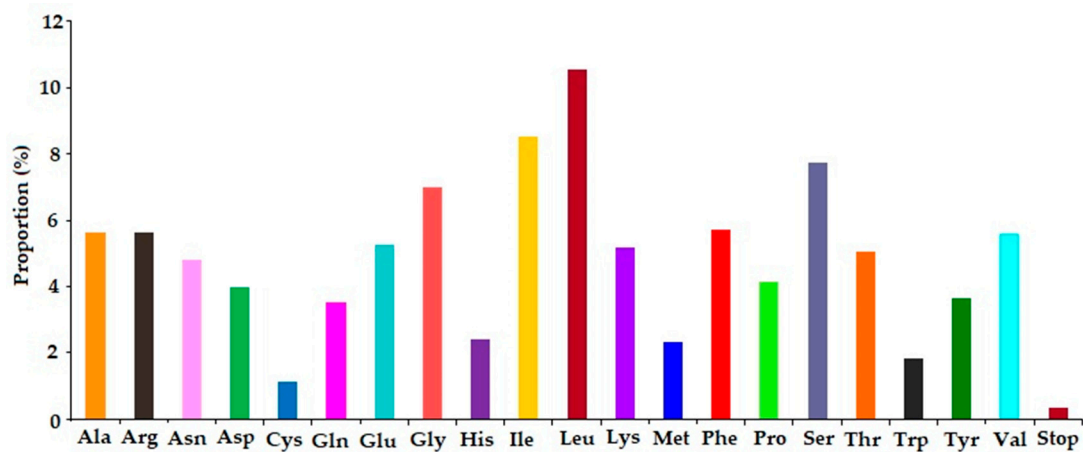


Figure 4. Amino acid frequencies in *A. altissima* chloroplast genome protein coding sequences.

Table 2. Predicted RNA editing site in the *A. altissima* chloroplast genome.

Gene	Nucleotide Position	Amino Acid Position	Codon Conversion	Amino Acid Conversion	Score
<i>accD</i>	818	273	T <u>C</u> G ≥ TTG	S ≥ L	0.80
<i>atpF</i>	92	31	C <u>C</u> A ≥ CTA	P ≥ L	0.86
	353	118	T <u>C</u> A ≥ TTA	S ≥ L	1.00
<i>atpB</i>	403	135	<u>C</u> CA ≥ TCA	P ≥ S	0.86
<i>rps14</i>	80	27	T <u>C</u> A ≥ TTA	S ≥ L	1.00
	149	50	T <u>C</u> A ≥ TTA	S ≥ L	1.00
<i>ccsA</i>	145	49	<u>C</u> TT ≥ TTT	L ≥ F	1.00
<i>clpP</i>	556	186	<u>C</u> AT ≥ TAT	H ≥ Y	1.00
	319	107	<u>C</u> TT ≥ TTT	L ≥ F	0.86
	457	153	<u>C</u> AC ≥ TAC	H ≥ Y	1.00
	643	215	<u>C</u> AT ≥ TAT	H ≥ Y	1.00
	1246	416	<u>C</u> AC ≥ TAC	H ≥ Y	1.00
<i>ndhA</i>	107	36	C <u>C</u> T ≥ CTT	P ≥ L	1.00
	341	114	T <u>C</u> A ≥ TTA	S ≥ L	1.00
	566	189	T <u>C</u> A ≥ TTA	S ≥ L	1.00
	1073	358	T <u>C</u> C ≥ TTC	S ≥ F	1.00
<i>ndhB</i>	149	50	T <u>C</u> A ≥ TTA	S ≥ L	1.00
	467	156	C <u>C</u> A ≥ CTA	P ≥ L	1.00
	586	196	<u>C</u> AT ≥ TAT	H ≥ Y	1.00
	611	204	T <u>C</u> A ≥ TTA	S ≥ L	0.80
	746	249	T <u>C</u> T ≥ TTT	S ≥ F	1.00
	830	277	T <u>C</u> A ≥ TTA	S ≥ L	1.00
	836	279	T <u>C</u> A ≥ TTA	S ≥ L	1.00
	1255	419	<u>C</u> AT ≥ TAT	H ≥ Y	1.00
1481	494	C <u>C</u> A ≥ CTA	P ≥ L	1.00	

Table 2. Cont.

Gene	Nucleotide Position	Amino Acid Position	Codon Conversion	Amino Acid Conversion	Score
<i>ndhD</i>	2	1	ACG ≥ ATG	T ≥ M	1.00
	313	105	CGG ≥ TGG	R ≥ W	0.80
	383	128	TCA ≥ TTA	S ≥ L	1.00
	674	225	TCA ≥ TTA	S ≥ L	1.00
	878	293	TCA ≥ TTA	S ≥ L	1.00
	887	296	CCT ≥ CTT	P ≥ L	1.00
	1076	359	GCT ≥ GTT	A ≥ V	1.00
	1298	433	TCA ≥ TTA	S ≥ L	0.80
	1310	437	TCA ≥ TTA	S ≥ L	0.80
<i>ndhF</i>	290	97	TCA ≥ TTA	S ≥ L	1.00
	586	196	CTT ≥ TTT	L ≥ F	0.80
	1919	640	GCT ≥ GTT	A ≥ V	0.80
<i>ndhG</i>	166	56	CAT ≥ TAT	H ≥ Y	0.80
	320	107	ACA ≥ ATA	T ≥ I	0.80
<i>petL</i>	119	40	CCT ≥ CTT	P ≥ L	0.86
<i>psbF</i>	77	26	TCT ≥ TTT	S ≥ F	1.00
<i>rpl20</i>	308	103	TCA ≥ TTA	S ≥ L	0.86
<i>rpoA</i>	830	277	TCA ≥ TTA	S ≥ L	1.00
<i>rpoB</i>	338	113	TCT ≥ TTT	S ≥ F	1.00
	551	184	TCA ≥ TTA	S ≥ L	1.00
	566	189	TCC ≥ TTG	S ≥ L	1.00
	2426	809	TCA ≥ TTA	S ≥ L	0.86
<i>rpoC1</i>	41	14	TCA ≥ TTA	S ≥ L	1.00
<i>rpoC2</i>	1681	561	CAT ≥ TAT	H ≥ Y	0.86
	2030	677	ACT ≥ ATT	T ≥ I	1.00
	2314	772	CGG ≥ TGG	R ≥ W	1.00
	4183	1395	CTT ≥ TTT	L ≥ F	0.80
<i>rps2</i>	248	83	TCA ≥ TTA	S ≥ L	1.00
<i>rps16</i>	209	70	TCA ≥ TTA	S ≥ L	0.83

The cytidines marked are putatively edited to uridines.

Comparisons of RNA editing sites with other six species from other families revealed that *R. chinensis* and *D. viscosa* have high RNA editing sites (61 each distributed in 20 and 17 genes respectively) followed by *B. sacra* (57 in 20 genes), *A. altissima* (54 in 21 genes), *A. indica* (53 in 21 genes), *C. aurantiifolia* (52 in 21 genes), and *L. floridana* 48 in 20 genes. As shown in Table S3, these results are consistent with several studies in that all the RNA editing sites predicted among the seven species are cytidine (C) to uridine (U) conversions [41,50–52]. Majority of RNA editing occurred at the second positions of the codons with a frequency from 62.30% (38/61) in *D. viscosa* to 81.28% (39/48) in *L. floridana*, which concurs with previous plastid genome studies in other land plants [53,54]. All the species shared 19 editing sites distributed in twelve genes (Table 3), whereas the two species from Simaroubaceae family (*L. floridana* and *A. altissima*) shared 33 editing sites in 16 genes this implies that the RNA editing sites in these two species are highly conserved (Table S4). Like previous studies [41,51,55], the *ndhB* gene in most of species analyzed here have the highest number of editing sites. Notably, a RNA editing event was detected at the initiator codon (ACG) resulting in ATG translational start codon in the *ndhD* gene.

2.4. Repeat Sequence Analysis

Microsatellites are usually 1–6 bp tandem repeat DNA sequences and are distributed throughout the genome. The presence of microsatellites was detected in the chloroplast genome of *A. altissima* (Figure 5). A total of 219 simple sequence repeats (SSRs) loci were detected, of which mononucleotide repeats occurred with high frequency constituting 190 (86.76%) of all the SSRs. Majority of mononucleotides composed of poly A (polyadenine) (39.27%) and poly T (polythymine) (47.49%) repeats, whereas poly G (polyguanine) or poly C (polycytosine) repeats were rather rare (2.74%). Among the dinucleotide repeat motifs AT/AT were more abundant, while AG/CT were less frequent. One trinucleotide motif (AAT/ATT), five tetra-(AAAG/CTTT, AAAT/ATTT, AACT/AGTT, AATC/ATTG and AGAT/ATCT) and two pentanucleotide repeats (AAAAG/CTTTT and AATAG/ATTCT) were identified. Hexanucleotide repeats were not detected in the *A. altissima* chloroplast genome.

As shown in Figure 5, the SSR analysis for seven species showed that *Leitneria floridana* had the highest number of SSRs (256) while *Dodonaea viscosa* and *Rhus chinensis* had the lowest (186). In all the seven species, mononucleotide repeats were more abundant with A/T repeats being the most common repeats. This result is consistent with earlier studies in [31,35,46] which revealed that many angiosperm chloroplast genomes are rich in poly A and poly T. Moreover, in the seven analyzed species, hexanucleotide repeats were not detected, whereas *Azadirachta indica*, *Dodonaea viscosa* and *Leitneria floridana* had no pentanucleotide repeats.

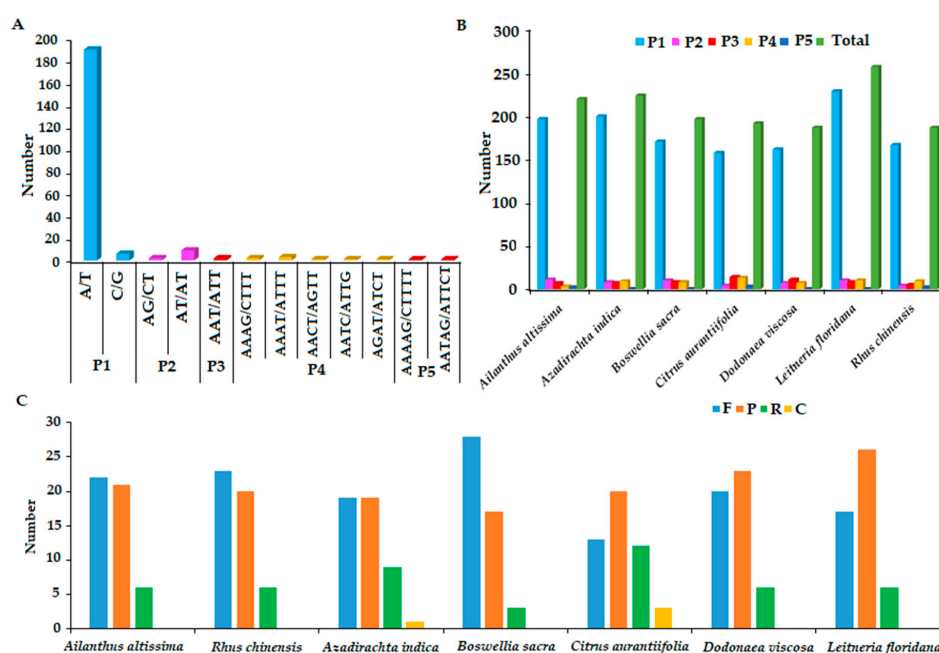


Figure 5. Simple sequence repeat (SSRs) type, distribution and presence in *A. altissima* and other representative species from Sapindales. (A) Number of detected SSR motifs in different repeat types in *A. altissima* Chloroplast genome. (B) Number of identified repeat sequences in seven chloroplast genomes. (C) Number of different SSR types in seven representative species. F, indicate (forward), P (palindromic), R (reverse), and C (complement), while P1, P2, P3, P4, P5 indicates Mono-, di-, tri-, tetra-, and penta-nucleotides respectively. F: forward; P: palindromic, R: reverse; C: complement.

The REPuter program revealed that *A. altissima* chloroplast genome contains 21 palindromic (p), 22 forward (f) and six reverse (r) repeats, however the complement repeats were not detected (Table 4). We notice that all the identified tandem repeats in *A. altissima* were more than 20 bp, while thirteen had length of more than 30 bp. Repeat analyses of seven Sapindales revealed a total of 48 or 49 repeats for each species, with all species containing forward, palindromic and reverse repeats

(Figure 5). Compliment repeats were not identified in other species except for *Azadirachta indica* and *Citrus aurantiifolia* which had one and three repeats respectively. *Citrus aurantiifolia* had the highest number of reverse repeats but also lowest number of forward repeats. Most of the repeat lengths were less than 50 bp, however *Boswellia sacra* chloroplast had seven forward repeats with length of between 65 to 251 bp. Besides, we found out that almost all the repeat sequences were located in either IR or LSC region.

Table 4. Distribution and localization of repetitive sequences F, forward; P, palindromic, R; reverse in *A. altissima* chloroplast genome.

Number	Size	Position 1	Type	Position 2	Location 1 (2)	Region
1	48	95,957	F	95,975	<i>ycf2</i>	IRa
2	48	153,174	F	153,192	<i>ycf2</i>	IRb
3	37	103,326	F	125,821	<i>rps12/trnV-GAC(ndhA*)</i>	IRa/SSC
4	30	95,957	F	95,993	<i>ycf2</i>	IRa
5	30	153,174	F	153,210	<i>ycf2</i>	IRb
6	29	50,944	F	50,972	<i>trnL-UAA*</i>	LSC
7	29	58,040	F	58,078	<i>rbcL</i>	LSC
8	28	115,434	F	115,460	<i>ycf1</i>	SSC
9	26	39,399	F	39,625	<i>psbZ/trnG-UCC</i>	LSC
10	25	71,153	F	71,178	<i>trnP-GGG/psaJ</i>	LSC
11	23	47,036	F	103,323	<i>ycf3**(rps12/trnV-GAC)</i>	LSC/IRa
12	23	112,456	F	112,488	<i>rrn4.5/rrn5</i>	IRa
13	23	136,686	F	136,718	<i>rrn5/rrn4.5</i>	IRb
14	22	11,749	F	11,771	<i>trnR-UCU/atpA</i>	LSC
15	21	248	F	270	<i>trnH-GUG/psbA</i>	LSC
16	21	9541	F	38,293	<i>trnS-GCU (trnS-UGA)</i>	LSC
17	21	41,956	F	44,180	<i>psaB(psaA)</i>	LSC
18	21	49,678	F	49,699	<i>trnL-UAA*</i>	LSC
19	20	1945	F	1965	<i>trnK-UUU</i>	LSC
20	20	15,166	F	92,503	<i>atpH/atpI(ycf2)</i>	LSC
21	20	47,039	F	125,821	<i>ycf3**(rps15)</i>	LSC/IRa
22	20	88,907	F	160,270	<i>rpl2</i>	IRa/IRb
25	48	31,790	P	31,790	<i>petN/psbM</i>	LSC
26	48	95,957	P	153,174	<i>ycf2</i>	IRa/IRb
27	48	95,975	P	153,192	<i>ycf2</i>	IRa/IRb
28	37	125,821	P	145,834	<i>ndhA*(trnV-GAC/rps12)</i>	SSC/IRb
29	36	30,970	P	30,970	<i>petN/psbM</i>	LSC
30	30	72,117	P	72,117	<i>rpl33/rps18</i>	LSC
31	30	95,957	P	153,174	<i>ycf2</i>	IRa/IRb
32	30	95,993	P	153,210	<i>ycf2</i>	IRa/IRb
33	27	542	P	571	<i>trnH-GUG/psbA</i>	LSC
34	25	11,403	P	11,430	<i>trnS-GCU/trnR-UCU</i>	LSC
35	24	4867	P	4897	<i>trnK-UUU/rps16</i>	LSC
36	24	9535	P	48,164	<i>trnS-GCU(psaA/ycf3)</i>	LSC
37	23	47,036	P	145,851	<i>ycf3**(trnV-GAC/rps12)</i>	LSC/IRb
38	23	51,804	P	119,066	<i>trnF-GAA/ndh](rpl32/trnL-UAG)</i>	LSC/SSC
39	23	112,456	P	136,686	<i>rrn4.5/rrn5</i>	IRa/IRb
40	23	112,488	P	136,718	<i>rrn4.5/rrn5</i>	IRa/IRb
41	22	39,195	P	39,195	<i>psbZ/trnG-UCC</i>	LSC
42	20	15,166	P	156,674	<i>atpH(ycf2)</i>	LSC/IRb
43	20	38,361	P	48,100	<i>trnS-UGA(trnS-GGA)</i>	LSC
44	20	88,907	P	88,907	<i>rpl2</i>	IRa
45	20	107,097	P	107,130	<i>rrn16/trnI-GAU</i>	IRa
46	23	39,184	R	39,184	<i>psbZ/trnG-UCC</i>	LSC
47	21	9751	R	9751	<i>trnS-GCU/trnR-UCU</i>	LSC
48	21	51,281	R	51,281	<i>trnL-UAA/trnF-GAA</i>	LSC
49	21	85,055	R	85,055	<i>rps8/rpl14</i>	LSC
50	20	53,712	R	53,712	<i>ndhC</i>	LSC
51	20	9385	R	13,356	<i>psbI(atpA/atpF)</i>	LSC

F: forward; P: palindrome; R; reverse* intron or ** introns.

2.5. Phylogenetic Tree

The phylogenetic position of *A. altissima* within Sapindales was carried out using 75 protein coding sequences shared by thirty-one taxa from Sapindales (Table S5). Three remaining species were from Thymelaeaceae family (*Aquilaria sinensis*) and Malvaceae (*Theobroma cacao* and *Abelmoschus esculentus*) from order Malvales selected as outgroups (Figure 6). The maximum likelihood (ML) analysis produced a phylogenetic tree which fully supported *A. altissima* to be closely related with *Leitneria floridana* with 100% bootstrap value. The ML resolved 26 nodes with high branch support (over 98% bootstrap values), however six nodes were moderately supported perhaps as a result of less samples use (59 to 95%). Concerning relationships among families within Sapindales order, family Simaroubaceae early diverged and formed a sister clade/relationship with a 95% bootstrap support to Rutaceae family. Interestingly, the placement of families within Sapindales in our phylogenetic tree supports the one reported by previous studies [30,56,57] based on some chloroplast and nuclear markers. The families Anacardiaceae and Burseraceae formed a sister clade/ group, this clade further branched forming a sister clade with families Sapindaceae, Meliaceae, Simaroubaceae and Rutaceae analyzed in our study. Therefore, it is crucial to use more species for better understanding of Simaroubaceae phylogeny and evolution. This study provides a basis for future phylogenetic of Simaroubaceae species.

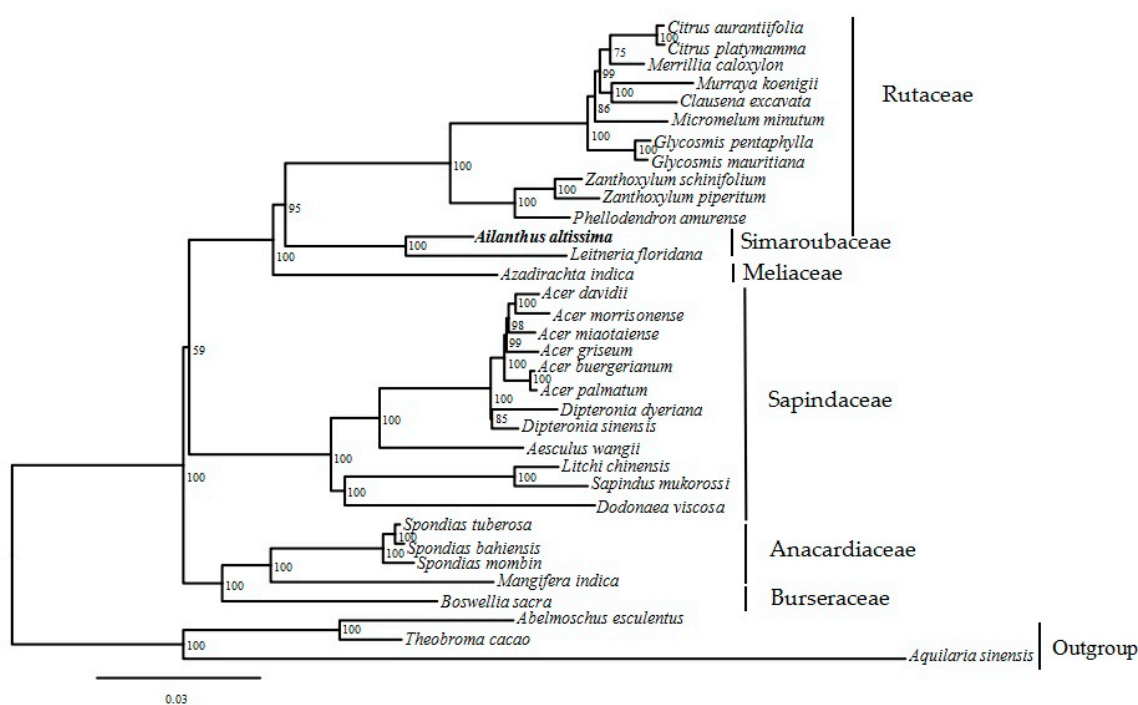


Figure 6. Phylogenetic tree of 31 Sapindales species with three outgroup Malvales species inferred from ML (Maximum likelihood) based on common protein coding genes. The position of *A. altissima* is shown in bold, while bootstrap support values are shown at each node.

3. Materials and Methods

3.1. Plant Materials and DNA Extraction

Fresh leaves of *Ailanthus altissima* were collected in Wuhan Botanical Garden, Chinese Academy of Sciences in China. Total genomic DNA isolation was carried out using MagicMag Genomic DNA Micro Kit (Sangon Biotech Co., Shanghai, China) based on the manufactures protocol. The quality and integrity of DNA were checked and inspected using spectrophotometry and agarose gel electrophoresis respectively. The voucher specimen (HIB-LZZ-CC003) has been deposited at the Wuhan Botanical Garden herbarium (HIB) Wuhan, China.

3.2. The Tree of Heaven Plastome Sequence Assembly and Annotation

Library preparation was constructed using the Illumina HiSeq 2500 platform at NOVogene Company (Beijing, China) with an average insert size of approximately 350 bp. The high-quality data (5 Gb) were filtered from raw sequence data (5.2 Gb) using the PRINSEQ lite v0.20.4 (San Diego State University, San Diego, CA, USA) [58] (phredQ \geq 20, Length \geq 50), followed by de novo assembling using NOVOPlasty [59] with default sets (K-mer = 31). The seeds and reference plastome used was from the closely related species *Leitneria floridana* (NC_030482) with high coverage of chloroplast reads $\sim 1500\times$. Lastly, one contig of *Ailanthus altissima* was generated and mapped with reference plastome using GENEIOUS 8.1 (Biomatters Ltd., Auckland, New Zealand) [60]. Finally, online web-based server local blast was used to verify the inverted repeat (IR) single copy (SC) junctions.

Preliminary gene annotation of assembled genome was done using the program DOGMA (Dual Organellar GenoMe Annotator, University of Texas at Austin, Austin, TX, USA) [61], and BLAST (<http://blast.ncbi.nlm.nih.gov/>). The positions of start and stop codons together with position of introns were confirmed by comparing with homologous genes of other chloroplast genomes available at the GenBank database. Moreover, tRNA genes were verified with tRNAscan-SE server (<http://lowelab.ucsc.edu/tRNAscan-SE/>) [62]. The chloroplast genome physical circular map was drawn using program OGDRAW (Organellar Genome DRAW) [63] Max planck Institute of Molecular Plant Physiology, Potsdam, Germany) accompanied by manual corrections. The chloroplast genome sequence of *A. altissima* was deposited in the GenBank database, accession (MG799542).

3.3. Genome Comparison and Gene Rearrangement

The border region between Inverted repeat (IR) and large single copy (LSC), also between inverted repeats and small single copy (SSC) junction were compared among seven representative species from Sapindales order. Additionally, alignments of seven chloroplast with one reference genome to determined gene rearrangements was carried out using Mauve v.4.0 [64].

3.4. Repeat Analysis in *A. altissima* Chloroplast Genome

Microsatellites were identified in the tree of heaven chloroplast genome and other selected representative genomes belonging to order Sapindales using an online software MICOSSATellite (MISA) [65]. The minimum number of repetitions were set to eight repeat units for mononucleotide SSR motifs, five repeat units for dinucleotide SSRs, four for trinucleotide SSRs and three repeat units for tetra-, penta-, and hexanucleotide motifs. The REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) program [66] (University of Bielefeld, Bielefeld, Germany) with default parameters was used to identify the location and sizes of forward, palindromic, complement and reverse repeat sequences in *A. altissima* chloroplast genome.

3.5. Codon Usage and RNA Editing Sites

CodonW1.4.2 (<http://downloads.fyxm.net/CodonW-76666.html>) was used to analyze codon usage. Subsequently, possible RNA editing sites in *A. altissima* protein coding genes were predicted using the program predictive RNA editor for plants (PREP) suite [67] with the cutoff value set to 0.8. PREP server uses 35 genes as reference for potential RNA editing sites prediction by comparing the predicted protein genes to homologous proteins from other plants.

3.6. Phylogenetic Analysis

Seventy five protein coding sequences present in 31 species from order Sapindales and three species from Thymelaeaceae (*Aquilaria sinensis*), Malvaceae (*Theobroma cacao* and *Abelmoschus esculentus*) as outgroups were used for the phylogenetic reconstruction. These species chloroplast genomes were downloaded from GenBank (Table S5). The protein coding sequences alignment was done using GENEIOUS v8.0.2 (Biomatters Ltd., Auckland, New Zealand) [60]. Maximum likelihood (ML) analysis

was carried out using RAxML version 8.0.20 (Scientific Computing Group, Heidelberg Institute for Theoretical Studies, Institute of Theoretical Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany) [68] with 1000 replicates for bootstrap test. Lastly, the jModelTest v2.1.7 [69] was used to select the best substitution model (GTR + I + G).

4. Conclusions

In this study, we present the plastome of tree of heaven from family Simaroubaceae which contains about 22 genera and over 100 species. *A. altissima* chloroplast genome possess circular and quadripartite structure which is well conserved similar to other plants chloroplast genomes. Nonetheless, the plastome showed slight variations at the four boundary junctions due to expansion and contraction in SC and IR borders. About 219 SSR loci and 49 repeats sequences were identified in *A. altissima* genome, this provides genetic information for designing DNA molecular markers for analyzing gene pool dynamics and genetic diversity of *A. altissima* natural populations aiming dispersal mechanism of this invasive tree. The phylogenetic analysis performed using 75 protein coding genes of 34 species available at the GenBank database, comprising 3 outgroup species from Malvales and 31 species representing families from order Sapindales. The two species from family Simaroubaceae formed a cluster and were group together with other families to form a single clade (Sapindales). In addition, the RNA editing analysis in *A. altissima* genome identified a total of 54 possible editing sites in 21 chloroplast genes with C-to-U transitions being the most. The availability of this chloroplast genome provides a tool to advance the study of evolution and invasion in *A. altissima* in order to address present evolutionary, ecological and genetic questions regarding this species.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/19/4/929/s1>.

Acknowledgments: This work was supported by the National Natural Scientific Foundation of China (Grant No. 31500457) and CAS-TWAS President's Fellowship for International PhD Students.

Author Contributions: Yi-Ying Liao conceived and designed the experiment; Josphat K. Saina, Zhi-Zhong Li and Andrew W. Gichira assembled plastome sequence and revised the manuscript. Josphat K. Saina performed the experiments, analyzed data and wrote the manuscript. Yi-Ying Liao and Zhi-Zhong Li collected the plant material. All the authors read and approved submission of the final manuscript.

Conflicts of Interest: The authors declare that they have no conflict of interest.

Abbreviations

SC	Single copy
LSC	Large single copy
SSC	Small single copy
IR	Inverted repeat

References

1. Kowarik, I.; Säumel, I. Biological flora of central Europe: *Ailanthus altissima* (Mill.) Swingle. *Perspect. Plant Ecol. Evol. Syst.* **2007**, *8*, 207–237. [CrossRef]
2. Liao, Y.Y.; Guo, Y.H.; Chen, J.M.; Wang, Q.F. Phylogeography of the widespread plant *Ailanthus altissima* (Simaroubaceae) in China indicated by three chloroplast DNA regions. *J. Syst. Evol.* **2014**, *52*, 175–185. [CrossRef]
3. Kurokouchi, H.; Saito, Y.; Ide, Y. Genetic structure of the introduced heaven tree (*Ailanthus altissima*) in Japan: Evidence for two distinct origins with limited admixture. *Botany* **2014**, *93*, 133–139. [CrossRef]
4. Dallas, J.F.; Leitch, M.J.; Hulme, P.E. Microsatellites for tree of heaven (*Ailanthus altissima*). *Mol. Ecol. Resour.* **2005**, *5*, 340–342. [CrossRef]
5. Aldrich, P.R.; Briguglio, J.S.; Kapadia, S.N.; Morner, M.U.; Rawal, A.; Kalra, P.; Huebner, C.D.; Greer, G.K. Genetic structure of the invasive tree *Ailanthus altissima* in eastern United States cities. *J. Bot.* **2010**, *2010*, 795735.
6. Neuhaus, H.; Emes, M. Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Biol.* **2000**, *51*, 111–140. [CrossRef] [PubMed]

7. Henry, R.J. *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*; CABI Publishing: Cambridge, MA, USA, 2005.
8. Raubeson, L.A.; Jansen, R.K. Chloroplast genomes of plants. In *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*; CABI Publishing: Cambridge, MA, USA, 2005; pp. 45–68.
9. Wicke, S.; Schneeweiss, G.M.; Müller, K.F.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297. [[CrossRef](#)] [[PubMed](#)]
10. Yue, F.; Cui, L.; Moret, B.M.; Tang, J. Gene rearrangement analysis and ancestral order inference from chloroplast genomes with inverted repeat. *BMC Genom.* **2008**, *9*, S25. [[CrossRef](#)] [[PubMed](#)]
11. Jansen, R.K.; Wojciechowski, M.F.; Sanniyasi, E.; Lee, S.-B.; Daniell, H. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of *rps12* and *clpP* intron losses among legumes (Leguminosae). *Mol. Phylog. Evol.* **2008**, *48*, 1204–1217. [[CrossRef](#)] [[PubMed](#)]
12. Lee, H.-L.; Jansen, R.K.; Chumley, T.W.; Kim, K.-J. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Mol. Biol. Evol.* **2007**, *24*, 1161–1180. [[CrossRef](#)] [[PubMed](#)]
13. Parks, M.; Cronn, R.; Liston, A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* **2009**, *7*, 84. [[CrossRef](#)] [[PubMed](#)]
14. Yi, X.; Gao, L.; Wang, B.; Su, Y.-J.; Wang, T. The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): Evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. *Genome Biol. Evol.* **2013**, *5*, 688–698. [[CrossRef](#)] [[PubMed](#)]
15. Moore, M.J.; Bell, C.D.; Soltis, P.S.; Soltis, D.E. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19363–19368. [[CrossRef](#)] [[PubMed](#)]
16. Wu, C.-S.; Wang, Y.-N.; Hsu, C.-Y.; Lin, C.-P.; Chaw, S.-M. Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and Cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. *Genome Biol. Evol.* **2011**, *3*, 1284–1295. [[CrossRef](#)] [[PubMed](#)]
17. Daniell, H.; Lin, C.-S.; Yu, M.; Chang, W.-J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 134. [[CrossRef](#)] [[PubMed](#)]
18. Jansen, R.K.; Saski, C.; Lee, S.-B.; Hansen, A.K.; Daniell, H. Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): Evidence for at least two independent transfers of *rpl22* to the nucleus. *Mol. Biol. Evol.* **2010**, *28*, 835–847. [[CrossRef](#)] [[PubMed](#)]
19. Cai, Z.; Penaflo, C.; Kuehl, J.V.; Leebens-Mack, J.; Carlson, J.E.; Boore, J.L.; Jansen, R.K. Complete plastid genome sequences of *drimys*, *Liriodendron*, and *Piper*: Implications for the phylogenetic relationships of magnoliids. *BMC Evol. Biol.* **2006**, *6*, 77. [[CrossRef](#)] [[PubMed](#)]
20. Steane, D.A. Complete nucleotide sequence of the chloroplast genome from the tasmanian blue gum, *Eucalyptus globulus* (myrtaceae). *DNA Res.* **2005**, *12*, 215–220. [[CrossRef](#)] [[PubMed](#)]
21. Wu, C.-S.; Lin, C.-P.; Hsu, C.-Y.; Wang, R.-J.; Chaw, S.-M. Comparative chloroplast genomes of Pinaceae: Insights into the mechanism of diversified genomic organizations. *Genome Biol. Evol.* **2011**, *3*, 309–319. [[CrossRef](#)] [[PubMed](#)]
22. Zalapa, J.E.; Cuevas, H.; Zhu, H.; Steffan, S.; Senalik, D.; Zeldin, E.; McCown, B.; Harbut, R.; Simon, P. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am. J. Bot.* **2012**, *99*, 193–208. [[CrossRef](#)] [[PubMed](#)]
23. Buschiazzo, E.; Gemell, N.J. The rise, fall and renaissance of microsatellites in eukaryotic genomes. *Bioessays* **2006**, *28*, 1040–1050. [[CrossRef](#)] [[PubMed](#)]
24. Kelkar, Y.D.; Tyekucheva, S.; Chiaromonte, F.; Makova, K.D. The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome Res.* **2008**, *18*, 30–38. [[CrossRef](#)] [[PubMed](#)]
25. Ebert, D.; Peakall, R. A new set of universal de novo sequencing primers for extensive coverage of noncoding chloroplast DNA: New opportunities for phylogenetic studies and CPSSR discovery. *Mol. Ecol. Resour.* **2009**, *9*, 777–783. [[CrossRef](#)] [[PubMed](#)]
26. Hirao, T.; Watanabe, A.; Miyamoto, N.; Takata, K. Development and characterization of chloroplast microsatellite markers for *Cryptomeria japonica* D. Don. *Mol. Ecol. Resour.* **2009**, *9*, 122–124. [[CrossRef](#)] [[PubMed](#)]

27. Do Nascimento Vieira, L.; Faoro, H.; Rogalski, M.; de Freitas Fraga, H.P.; Cardoso, R.L.A.; de Souza, E.M.; de Oliveira Pedrosa, F.; Nodari, R.O.; Guerra, M.P. The complete chloroplast genome sequence of *Podocarpus lambertii*: Genome structure, evolutionary aspects, gene content and SSR detection. *PLoS ONE* **2014**, *9*, e90618.
28. Yao, X.; Tang, P.; Li, Z.; Li, D.; Liu, Y.; Huang, H. The first complete chloroplast genome sequences in Actinidiaceae: Genome structure and comparative analysis. *PLoS ONE* **2015**, *10*, e0129347. [[CrossRef](#)] [[PubMed](#)]
29. Su, H.-J.; Hogenhout, S.A.; Al-Sadi, A.M.; Kuo, C.-H. Complete chloroplast genome sequence of Omani lime (*Citrus aurantiifolia*) and comparative analysis within the rosids. *PLoS ONE* **2014**, *9*, e113049. [[CrossRef](#)] [[PubMed](#)]
30. Lee, M.; Park, J.; Lee, H.; Sohn, S.-H.; Lee, J. Complete chloroplast genomic sequence of *Citrus platyamma* determined by combined analysis of sanger and NGS data. *Hortic. Environ. Biotechnol.* **2015**, *56*, 704–711. [[CrossRef](#)]
31. Saina, J.K.; Gichira, A.W.; Li, Z.-Z.; Hu, G.-W.; Wang, Q.-F.; Liao, K. The complete chloroplast genome sequence of *Dodonaea viscosa*: Comparative and phylogenetic analyses. *Genetica* **2018**, *146*, 101–113. [[CrossRef](#)] [[PubMed](#)]
32. Raman, G.; Park, S. The complete chloroplast genome sequence of *Ampelopsis*: Gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. *Front. Plant Sci.* **2016**, *7*, 341. [[CrossRef](#)] [[PubMed](#)]
33. Park, I.; Kim, W.J.; Yeo, S.-M.; Choi, G.; Kang, Y.-M.; Piao, R.; Moon, B.C. The complete chloroplast genome sequences of *Fritillaria ussuriensis* maxim. In addition, *Fritillaria cirrhosa* D. Don, and comparative analysis with other *Fritillaria* species. *Molecules* **2017**, *22*, 982. [[CrossRef](#)] [[PubMed](#)]
34. Li, B.; Lin, F.; Huang, P.; Guo, W.; Zheng, Y. Complete chloroplast genome sequence of *Decaisnea insignis*: Genome organization, genomic resources and comparative analysis. *Sci. Rep.* **2017**, *7*, 10073. [[CrossRef](#)] [[PubMed](#)]
35. Zhou, T.; Chen, C.; Wei, Y.; Chang, Y.; Bai, G.; Li, Z.; Kanwal, N.; Zhao, G. Comparative transcriptome and chloroplast genome analyses of two related *Dipteronia* species. *Front. Plant Sci.* **2016**, *7*, 1512. [[CrossRef](#)] [[PubMed](#)]
36. Kim, K.-J.; Lee, H.-L. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Res.* **2004**, *11*, 247–261. [[CrossRef](#)] [[PubMed](#)]
37. Wang, R.-J.; Cheng, C.-L.; Chang, C.-C.; Wu, C.-L.; Su, T.-M.; Chaw, S.-M. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol. Biol.* **2008**, *8*, 36. [[CrossRef](#)] [[PubMed](#)]
38. Wang, W.; Messing, J. High-throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA. *PLoS ONE* **2011**, *6*, e24670. [[CrossRef](#)] [[PubMed](#)]
39. Khan, A.L.; Al-Harrasi, A.; Asaf, S.; Park, C.E.; Park, G.-S.; Khan, A.R.; Lee, I.-J.; Al-Rawahi, A.; Shin, J.-H. The first chloroplast genome sequence of *Boswellia sacra*, a resin-producing plant in Oman. *PLoS ONE* **2017**, *12*, e0169794. [[CrossRef](#)] [[PubMed](#)]
40. Yang, J.; Yue, M.; Niu, C.; Ma, X.-F.; Li, Z.-H. Comparative analysis of the complete chloroplast genome of four endangered herbals of *Notopterygium*. *Genes* **2017**, *8*, 124. [[CrossRef](#)] [[PubMed](#)]
41. Wang, W.; Yu, H.; Wang, J.; Lei, W.; Gao, J.; Qiu, X.; Wang, J. The complete chloroplast genome sequences of the medicinal plant *Forsythia suspensa* (oleaceae). *Int. J. Mol. Sci.* **2017**, *18*, 2288. [[CrossRef](#)] [[PubMed](#)]
42. Yang, Y.; Zhu, J.; Feng, L.; Zhou, T.; Bai, G.; Yang, J.; Zhao, G. Plastid genome comparative and phylogenetic analyses of the key genera in Fagaceae: Highlighting the effect of codon composition bias in phylogenetic inference. *Front. Plant Sci.* **2018**, *9*, 82. [[CrossRef](#)] [[PubMed](#)]
43. Guo, S.; Guo, L.; Zhao, W.; Xu, J.; Li, Y.; Zhang, X.; Shen, X.; Wu, M.; Hou, X. Complete chloroplast genome sequence and phylogenetic analysis of *Paeonia ostii*. *Molecules* **2018**, *23*, 246. [[CrossRef](#)] [[PubMed](#)]
44. Shen, X.; Wu, M.; Liao, B.; Liu, Z.; Bai, R.; Xiao, S.; Li, X.; Zhang, B.; Xu, J.; Chen, S. Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant *Artemisia annua*. *Molecules* **2017**, *22*, 1330. [[CrossRef](#)] [[PubMed](#)]
45. Li, Z.-Z.; Saina, J.K.; Gichira, A.W.; Kyalo, C.M.; Wang, Q.-F.; Chen, J.-M. Comparative genomics of the Balsaminaceae sister genera *Hydrocera triflora* and *Impatiens pinfanensis*. *Int. J. Mol. Sci.* **2018**, *19*, 319. [[CrossRef](#)] [[PubMed](#)]

46. Zhou, J.; Chen, X.; Cui, Y.; Sun, W.; Li, Y.; Wang, Y.; Song, J.; Yao, H. Molecular structure and phylogenetic analyses of complete chloroplast genomes of two *Aristolochia* medicinal species. *Int. J. Mol. Sci.* **2017**, *18*, 1839. [[CrossRef](#)] [[PubMed](#)]
47. Jiang, D.; Zhao, Z.; Zhang, T.; Zhong, W.; Liu, C.; Yuan, Q.; Huang, L. The chloroplast genome sequence of *Scutellaria baicalensis* provides insight into intraspecific and interspecific chloroplast genome diversity in *Scutellaria*. *Genes* **2017**, *8*, 227. [[CrossRef](#)] [[PubMed](#)]
48. Zhou, J.; Cui, Y.; Chen, X.; Li, Y.; Xu, Z.; Duan, B.; Li, Y.; Song, J.; Yao, H. Complete chloroplast genomes of *Papaver rhoas* and *Papaver orientale*: Molecular structures, comparative analysis, and phylogenetic analysis. *Molecules* **2018**, *23*, 437. [[CrossRef](#)] [[PubMed](#)]
49. Chen, C.; Zheng, Y.; Liu, S.; Zhong, Y.; Wu, Y.; Li, J.; Xu, L.-A.; Xu, M. The complete chloroplast genome of *Cinnamomum camphora* and its comparison with related *Lauraceae* species. *PeerJ* **2017**, *5*, e3820. [[CrossRef](#)] [[PubMed](#)]
50. De Santana Lopes, A.; Pacheco, T.G.; Nimz, T.; do Nascimento Vieira, L.; Guerra, M.P.; Nodari, R.O.; de Souza, E.M.; de Oliveira Pedrosa, F.; Rogalski, M. The complete plastome of macaw palm [*Acrocomia aculeata* (jacq.) lodd. Ex mart.] and extensive molecular analyses of the evolution of plastid genes in arecaceae. *Planta* **2018**, 1–20. [[CrossRef](#)] [[PubMed](#)]
51. Kumbhar, F.; Nie, X.; Xing, G.; Zhao, X.; Lin, Y.; Wang, S.; Weining, S. Identification and characterisation of rna editing sites in chloroplast transcripts of einkorn wheat (*Triticum monococcum*). *Ann. Appl. Biol.* **2018**, *172*, 197–207. [[CrossRef](#)]
52. Huang, Y.-Y.; Cho, S.-T.; Haryono, M.; Kuo, C.-H. Complete chloroplast genome sequence of common Bermuda grass (*Cynodon dactylon* (L.) pers.) and comparative analysis within the family poaceae. *PLoS ONE* **2017**, *12*, e0179055.
53. Park, M.; Park, H.; Lee, H.; Lee, B.-H.; Lee, J. The complete plastome sequence of an antarctic bryophyte *Sanionia uncinata* (Hedw.) loeske. *Int. J. Mol. Sci.* **2018**, *19*, 709. [[CrossRef](#)] [[PubMed](#)]
54. Chen, H.; Deng, L.; Jiang, Y.; Lu, P.; Yu, J. RNA editing sites exist in protein-coding genes in the chloroplast genome of *Cycas taitungensis*. *J. Integr. Plant Biol.* **2011**, *53*, 961–970. [[CrossRef](#)] [[PubMed](#)]
55. De Santana Lopes, A.; Pacheco, T.G.; dos Santos, K.G.; do Nascimento Vieira, L.; Guerra, M.P.; Nodari, R.O.; de Souza, E.M.; de Oliveira Pedrosa, F.; Rogalski, M. The *Linum usitatissimum* L. Plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within malpighiales. *Plant Cell Rep.* **2018**, *37*, 307–328. [[CrossRef](#)] [[PubMed](#)]
56. Clayton, J.W.; Fernando, E.S.; Soltis, P.S.; Soltis, D.E. Molecular phylogeny of the tree-of-heaven family (Simaroubaceae) based on chloroplast and nuclear markers. *Int. J. Plant Sci.* **2007**, *168*, 1325–1339. [[CrossRef](#)]
57. Lee, Y.S.; Kim, I.; Kim, J.-K.; Park, J.Y.; Joh, H.J.; Park, H.-S.; Lee, H.O.; Lee, S.-C.; Hur, Y.-J.; Yang, T.-J. The complete chloroplast genome sequence of *Rhus chinensis* mill (Anacardiaceae). *Mitochondrial DNA Part B* **2016**, *1*, 696–697. [[CrossRef](#)]
58. Schmieder, R.; Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **2011**, *27*, 863–864. [[CrossRef](#)] [[PubMed](#)]
59. Dierckxsens, N.; Mardulyn, P.; Smits, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **2016**, *45*, e18.
60. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [[CrossRef](#)] [[PubMed](#)]
61. Wyman, S.K.; Jansen, R.K.; Boore, J.L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **2004**, *20*, 3252–3255. [[CrossRef](#)] [[PubMed](#)]
62. Schattner, P.; Brooks, A.N.; Lowe, T.M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **2005**, *33*, W686–W689. [[CrossRef](#)] [[PubMed](#)]
63. Lohse, M.; Drechsel, O.; Bock, R. OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **2007**, *52*, 267–274. [[CrossRef](#)] [[PubMed](#)]
64. Darling, A.C.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [[CrossRef](#)] [[PubMed](#)]

65. Thiel, T.; Michalek, W.; Varshney, R.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422. [[CrossRef](#)] [[PubMed](#)]
66. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. Reputer: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [[CrossRef](#)] [[PubMed](#)]
67. Mower, J.P. The prep suite: Predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* **2009**, *37*, W253–W259. [[CrossRef](#)] [[PubMed](#)]
68. Stamatakis, A. Raxml version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)] [[PubMed](#)]
69. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).