Computational and Structural Biotechnology Journal 18 (2020) 2424–2437

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Large inverted repeats identified by intra-specific comparison of mitochondrial genomes provide insights into the evolution of *Agrocybe aegerita*

Xinrui Liu [a], Xiaoping Wu [a], Hao Tan [b,c,*], Baogui Xie [a,*], Youjin Deng [a,*]

[a] College of Life Science, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China
[b] Mushroom Research Center, Soil and Fertilizer Institute, Sichuan Academy of Agricultural Sciences, Chengdu 610000, China
[c] School of Bioengineering, Jiangnan University, Wuxi 214062, China

ARTICLE INFO

Article history:
Received 19 May 2020
Received in revised form 25 August 2020
Accepted 26 August 2020
Available online 2 September 2020

Keywords:
Mitochondrial genotype
Linear plasmid
Short inverted repeat
Short insertion

ABSTRACT

Genomic structure and content of *Agrocybe aegerita* mitochondrial DNA contain essential information regarding the evolution of this gourmet mushroom. In this study, eight isolates of *A. aegerita* were sequenced and assembled into complete mitochondrial genomes. The mtDNA of the isolate Ag0067 contained two genotypes, both of which were quadripartite architecture consisting of two identical inverted repeats, separated by a small single-copy region and a large single-copy region. The only difference was opposite directions of the small single-copy region. The mtDNAs ranged from 116,329 bp to 134,035 bp, harboring two large identical inverted repeats. Genes of plasmid-origin were present in regions flanked by inverted repeat ID2. Most of the core genes evolved at a relatively low rate, whereas five tRNA genes located in corresponding regions of Ag0002:1–14000 and Ag0002:50001–61000 showed higher diversity. A long fragment inversion (10 Kb) was suggested to have occurred during the differentiation of two main clades, leading to two different gene orders. The number and distribution of the introns varied greatly among the *A. aegerita* mtDNAs. Fast invasion of short insertions likely resulted in the diversity of introns as well as other non-coding regions, increasing the variation of the mtDNAs. We raised a model about the evolution of the large repeats to explain the unusual features of *A. aegerita* mtDNAs. This study constructed quadripartite architecture of *A. aegerita* mtDNAs analogous to chloroplast DNA, proposed an interconversion model of the divergent mitochondrial genotypes with large inverted repeats. The findings could increase our knowledge of fungal evolution.

1. Introduction

Mitochondrion is an important cellular organelle in almost all eukaryotic organisms responsible for producing adenosine triphosphate (ATP). It is also involved in cellular processes of iron balance maintenance, cell senescence during cell cycle, apoptosis, and etc [1,2]. Mitochondrial DNAs are generally considered to have separate evolutionary origin from nuclear ones, which are thought to derive from the circular genome of a vestigial endosymbiotic proteobacterial ancestor [3]. Most of the genes have been transferred to nuclear genome during long-term evolution, resulting in reduced size and gene content of mitochondria DNAs than those in free-living proteobacteria [3].

With a rapid development of high throughput-sequencing technologies, an increasing number of mtDNAs in filamentous fungi have been released recently. Generally, they contain highly conserved core genes, including fourteen protein coding genes (PCGs), two rRNA genes, and a set of tRNA genes, all of which are responsible for partial functions of their hosts. The mitochondrial genomes vary greatly in length, ranging from 12 Kb (*Rozella allomycis* [4]) to 236 Kb (*Rhizoctonia solani* Ag-3 isolate [5]). Some mobile elements such as introns, linear plasmids, short repeat sequences, can be highly variable in number and length, leading to size polymorphism of mitochondrial genomes [6–8].

Core genes in fungal mtDNAs often possess large introns longer than 150 bp, which contain sequences with conserved RNA

* Corresponding authors at: College of Life Science, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China (Y. Deng & B. Xie); Mushroom Research Center, Soil and Fertilizer Institute, Sichuan Academy of Agricultural Sciences, Chengdu 610000, China, and School of Bioengineering, Jiangnan University, Wuxi 214062, China (H. Tan).

E-mail addresses: h.tan@foxmail.com (H. Tan), mrcfafu@163.com (B. Xie), dengyoujin1980@163.com (Y. Deng).

https://doi.org/10.1016/j.csbj.2020.08.022

secondary structures involved in autocatalytic splicing [9]. According to RNA secondary structures, the introns are classified into two major groups [9]. Group I introns, predominant in fungal mitochondrial genomes [10], are further clustered into several subgroups [9]. ORFs in group I introns are usually homing endonuclease genes (HEGs) functioning for transfer and site-specific integration of their host introns. Two families of HEGs, LAGLIDADG and GIY-YIG, are generally found in fungal group I introns. HEG is a main element responsible for host-intron movement. Destruction of HEG may hinder the insertion and/or deletion of introns [11]. Contrary with Group I, Group II introns rarely present in fungal mtDNAs.

Linear plasmids are another kind of mobile elements present frequently in fungal mtDNAs [12]. Most of the plasmids possess two large ORFs with well-known functions. One of the ORFs encodes a family-B DNA polymerase, and the other encodes a subunit of RNA polymerase [13]. Having little or no homology with mtDNA, the plasmids were thought to have a distinct evolutionary path from their host mtDNAs [14]. In Basidiomycota, the mitochondria-related plasmids have been reported to be free from mtDNAs in a number of species such as *Flammulina velutipes* [15] and *Pleurotus ostreatus* [16]. These mobile elements can also integrate into the mitochondrial genomes, for example, of *Agaricus bisporus* [13], *Agrocybe aegerita* [17,18], *P. ostreatus* [19] and *Moniliophthora perniciosa* [14].

Besides these mobile elements with large size, insertion or deletion of short segments, such as short inverted repeats (SIRs) also contributed to the diversity of intra-specific mtDNAs [20]. Short inverted repeats contain nucleotide sequences able to form hairpin structures, which provide protein-interacting sites. It facilitates the involvements of SIRs in cellular processes, such as DNA replication, transcriptional regulation, and recombination [21,22]. SIRs have been identified in mtDNAs of all eukaryotic organisms, with the highest variability in fungi and plants, and the lowest variability in mammals [23]. They are not distributed randomly, and are enriched in replication-originated sequences, D-loop, stem-loop, and miscellaneous [23]. Fungal mtDNA was homoplasmic, small in size, with high copy number and low level of recombination. These features make it suitable for distinguishing closely related species or even intra-specific isolates, as well as deduction of recent evolutionary events. Intra-specific mtDNA comparison has been used in fungal species such as *Cordyceps militaris* [24], Schizosaccharomyces pombe [25], *Isaria cicadae* [10], and *Colletotrichum lindemuthianum* [26]. These studies provided new insights into population evolution, identification of highly polymorphic regions, and isolate-specific markers [24].

A. aegerita (Brig.) Singer, also called *Agrocybe cylindracea* (DC.: Fr) Singer, is a medicinally and nutritionally important edible basidiomycete cultivated in numerous Asian and European countries [27–30] (Fig. 1). It is regarded as a multispecies complex, due to unsuccessful cross-breeding among some isolates as well as great variation of its morphological and physiological characters [31]. *A. aegerita* has been studied as a model mushroom of mitochondrial genes and genomes for the past three decades. An *A. aegerita* mitochondrial DNA of 80,500 bp length was assessed based on a restriction map in 1992 [32]. The mtDNA possesses two *polB* of linear-plasmid origin [17], which is flanked by two large inverted repeats longer than 2421 bp [33]. Each repeat contained an identical copy of *nad4* gene. Two polymorphic regions (the region around *cox2* gene, and the region carrying *cox1*, *atp6*, and *atp8* genes) within the mtDNA were reported to vary independently. Genome rearrangement is one of the major factors driving the polymorphism [34]. A complete *A. aegerita* mtDNA of 116,329 bp was reported in 2017 by Xu LM, et al. [35], that briefly described the structure and key features of the mtDNA.

In this study, we sequenced the mitochondrial genomes of *A. aegerita* Ag0067 using PacBio sequencing, and seven other isolates using Illumina sequencing platform. Based on long PacBio reads, mtDNA organization including large inverted repeats in Ag0067 was investigated. The eight complete mtDNAs as well as a relative isolate SWS_17 (SWS17) [35] released before were compared to provide a comprehensive picture of intra-specific mtDNA sequence polymorphism in *A. aegerita*. Our analyses depicted the intra-specific genic, intronic, and intergenic sequence structure in *A. aegerita* mitochondrial genomes, which might reveal a potential way of large inverted repeats evolution.

2. Materials and methods

2.1. Sample collection and DNA extraction

Eight isolates of *A. aegerita* were used in this study (Supplementary Table 1). Briefly, Ag0002 and Ag0067 were monokaryotic isolates obtained through protoplast monokaryolization from their parental isolates, whereas the other six isolates Ag0033, Ag0058, Ag0065, Ag0072, Ag0085, and Ag0086 were dikaryotic isolates. Isolates Ag0002 is a white mutant donated by Taiwan Food Industry Research and Development Institute to Sanming Fungi Research Institute in 1999. Isolate Ag0033 with brown fruiting bodies was donated by Shanghai Mushroom Research Institute to Sanming Fungi Research Institute in 1998, whose source was collected from Taiwan. Professor Meiying Guo of Sanming Fungi Research Institute donated the collections to Fujian Agriculture and Forestry University in 2009. The other six were wild isolates from different regions of Fujian. The eight isolates, including Ag0002 and Ag0033, are now all preserved in the Edible Fungal Germplasm Resources Management Center of Fujian Province, Fuzhou, China. Vegetative mycelium was grown in potato dextrose broth (PDB) medium at

Fig. 1. A. Wild *A. aegerita* grew in association with tea trees. B. Artificially cultivated *A. aegerita*.

25 ℃ for six days with 100 rpm shaking. Total genomic DNA for Illumina sequencing was extracted using a PH plant DNA kit (D2485-200, Omega Bio-Tek) following the procedures described in the manufacturer's instructions. DNA extraction and genome sequencing were performed by Novogene Corporation Inc (Beijing, China).

2.2. Mitochondrial genome sequencing and assembly

Genomic DNA of the Ag0067 isolate was sequenced using a PacBio Sequel system, thereby obtaining approximately 100 × long reads. These reads were assembled into genomic sequences by the program Canu 1.8 [36] with default parameters. Mitochondrial-related contigs were picked up from the assembly by BLASTn using the mitochondrial sequence of *A. aegerita* cultivar SWS_17 (SWS17, Accession No. MF979820) as a template. mtDNAs of *A. aegerita* were assumed to be circular, like most of the published fungal mitochondrial genomes. These contigs were linked into circular scaffolds using long PacBio reads. All reads were then mapped back to confirm the assemblies. The assembled mitochondrial sequences were polished using corresponding Illumina reads by the Pilon-1.21 program [37]. Genomic DNAs of the other seven isolates were sequenced using Illumina HiSeq X Ten, generated 4 Gb raw data for each. The Illumina reads were assembled into contigs by SPAdes-3.7.1 [38] with K-mer 127. Mitochondrial-related contigs were picked up from the assemblies based on sequencing coverage (more than three times of the average coverage in nuclear genome) as well as the framework of the mitochondrial DNA (mtDNA) of Ag0067. The picked contigs were linked with each other according to the pair-end relationship of Illumina reads. mtDNA of SWS17 was retrieved from database of NCBI for further analyses of intra-specific comparison of mitochondrial genome.

Repeat sequences within mtDNA were identified by BLASTn of mtDNA against itself with an E-value threshold of $1.00 \times e^{-5}$. The Illumina sequencing depth of each site in mtDNA was counted by a combination of the bwa 0.7.17 [39] and samtools 1.7 programs [40], to determine the presence of large repeats. Assembly of the large repeats was confirmed by aligning with long PacBio reads. Alignments of mitochondrial genome with reads were shown by the program IBS 1.0 [41].

2.3. Mitochondrial genome annotation

The assembled mitochondrial genomes of *A. aegerita* isolates were annotated mainly by online tools MFannot (http://megasun.bch.umontreal.ca/cgi-bin/dev_mfa/mfannotInterface.pl) according to the mould mitochondrial genetic code. Mini introns within *rns* and *rnl* genes were predicted manually by BLASTn using intron-free gene as query against genes with intron. Intron-exon boundaries of protein-coding genes were adjusted manually based on BLASTn against intron-free genes at corresponding positions. Conserved domains within the introns were predicted by searching in the conserved domain database (CDD) in NCBI. Mitochondrial genomic map was generated with the shinyCircos software [42].

2.4. Alignment and phylogenetic analysis

Mitochondrial coding sequences were extracted from the nine *A. aegerita* isolates. Multiple sequence alignment was performed by Clustal W [43] as implemented in MEGA-X [44]. Estimation of π, polymorphic sites was calculated using DnaSP v5 [45].

The concatenated coding sequence (CDS) of 14 conserved protein-coding genes (atp6-atp8 -atp9-cob-cox1-cox2-cox3-nad1-nad2-nad3-nad4-nad4L-nad5-nad6) was used to construct a phylogenetic tree using *Crinipellis perniciosa* (Accession No. AY376688) as outgroup. Pairwise and multiple alignments were performed by Clustal W [43] with a gap-opening penalty of 15 and a gap-extension penalty of 6.66. The phylogenetic tree was constructed by MEGA-X [44] with the Neighbor-Joining method. Gaps or missing data were deleted completely. A bootstrap method was used to test the phylogenetic tree with 1000 replications. Evolutionary distance was calculated using the Maximum Composite Likelihood method.

In order to analyze the evolution of plasmid in the tested mtDNAs, phylogenetic trees of B-type DNA polymerase genes and small orthologs of linear-plasmid origin were constructed respectively based on their amino-acid sequences. Methods of sequence alignment and tree construction were the same as those for the concatenated CDS of the conserved protein-coding genes.

2.5. Identification of short insertions/deletions

Comparison of intra-specific mtDNAs revealed a lot of short insertions/deletions (SI/Ds, <200 bp), many of which were multiple in corresponding genomes. The SI/Ds were detected by pairwise alignment of their mitochondrial sequences using program Clustal W [46]. Copy number was counted by BLASTn against mtDNAs using the SI/Ds as queries with 10% mismatches allowed. The SI/Ds were divided into clusters using Usearch [47] at an identity threshold of 90%. Each group of the fragments was regarded as a type of SI/D. SI/Ds with similar sequences but different lengths were clustered using CD-HIT [48] at an identity threshold of 0.9. Number of palindromes within each SI/D was counted by online tool palindrome analyser [49]. Palindrome size was set from 6 to 30 bp, spacer size 0 to 10 and maximally one mismatch was allowed.

3. Results

3.1. Mitochondrial genome assembly

PacBio Sequel was used to obtain a complete mitochondrial (mt) DNA of Ag0067. A total of 0.49 million reads with an average length of 10,388 bp were generated, accumulating up to a raw data of 5.1 Gb. These reads were assembled to genomic sequences of 54.5-Mb, among which 16 contigs belonged to a mitochondrial genome. Using PacBio reads longer than 36 Kb, the contigs were assembled into two circular mtDNA genotypes, type I and type II (Fig. 2B). Both genotypes were 121,484 bp in size (Accession No. MT364879), and had a quadripartite structure consisting of two identical inverted repeats (ID2 and ID4) of 31,173 bp, divided by a small single-copy region (ID1, 15,068 bp) and a large single-copy region (ID4, 44,071 bp). The only difference between these two genotypes was the opposite directions of their ID1 regions (Fig. 2B). In order to confirm the assemblies of both genotypes, large PacBio reads were mapped back (Fig. 2C and D respectively). For type I, thirteen reads, including R002, R005, R014 and R018, covered the ID2 region; four reads, including R021, covered the ID4 region. For type II, three reads, including R028, R029, and R030, covered the ID2 region; four reads, including R024, R025, R026, and R027, covered the ID4 region. The frequency of type I and type II were 0.71 and 0.29, respectively. Using Illumina reads to map on the regions ID1-ID2-ID3, sequencing depth of ID2 region was twice of ID1 and ID3 regions (Fig. 2E). It confirmed that ID2 was repetitive in Ag0067 genome.

Whole genome sequencing using the Illumina platform was performed to assemble the mtDNAs of *A. aegerita* isolates Ag0002, Ag0033, Ag0058, Ag0065, Ag0072, Ag0085 and Ag0086 (Accession No. MT364880 - MT364886, respectively). A total of 2.0–7.4 Gb paired-end reads with 150 bp × 2 were obtained. The assembly generated a 48.0 Mb genomic sequence of a

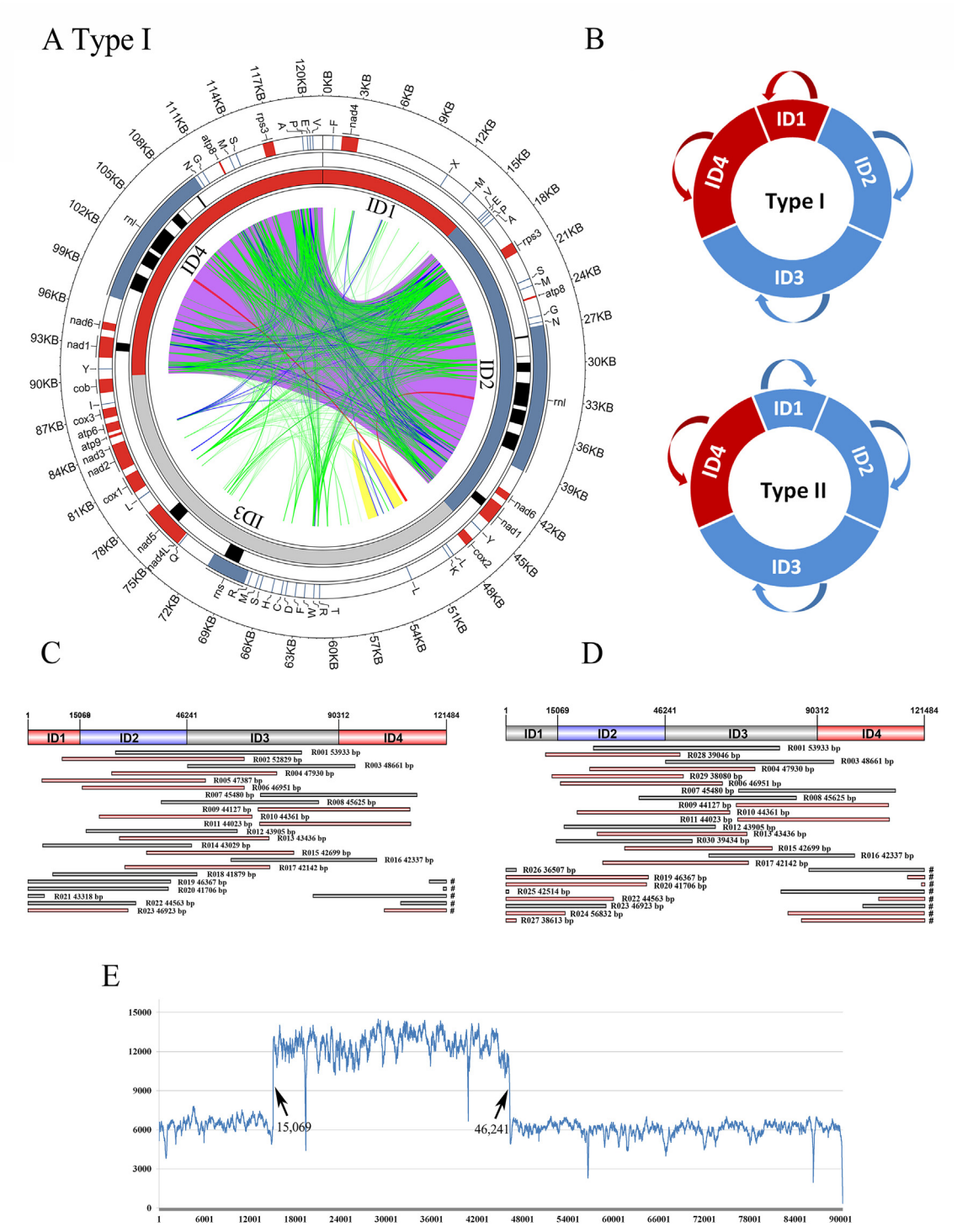


Fig. 2. Structures of the two mitochondrial genotypes in Ag0067. A, schematic diagram of the genotype I (Type I) in the Ag0067 mitochondrial genome. From outer to inner: genome position, gene order (red, protein-coding gene; blue, tRNA or rRNA gene), introns (black), regions (ID1-ID4), and linkage of the repeat sequences. Capital letters in circle of gene order, abbreviation of tRNA genes. B, Orientation of ID1, ID2, ID3, ID4 in two genotypes (Type I and Type II) of Ag0067 mitochondrial genome. C and D, Alignments of long PacBio reads with the Type I and Type II of the Ag0067 mtDNA. PacBio reads longer than 36 Kb were used for alignment. Gray reads, reads with clockwise alignment; pink reads, reads with counter-clockwise alignment; #, corresponding reads mapping onto both ends of the genotypes. E, Illumina sequencing depth of the regions ID1-ID3 in the Ag0067 genotype I mtDNA. Arrows point to jumping sites of sequencing depth. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

monokaryotic isolate Ag0002, as well as dikaryotic isolates between 62.2 and 78.4 Mb (Supplementary Table 1). From the genomes, 4–18 contigs were detected to be mitochondrial-related. Using the type I mtDNA of Ag0067 as a reference, the mitochondrial-related contigs as well as corresponding reads were assembled into a circular mtDNA for each isolate, respectively. Each assembly contained a large inverted repeat. The sequence and the length of the large inverted repeat were in line with the results of analyzing the sequencing depth (Supplementary Fig. 1). mtDNAs of the seven isolates ranged from 123,386 bp to

134,035 bp in size, and possessed a large inverted repeat, measuring between 28,899 bp and 35,145 bp (Table 1). The previously released mitochondrial genome of SWS17 had a 116,329 bp section corresponding to type II mtDNA. The section contained sequences corresponding to the ID1 to ID4 regions of Ag0067. The inverted repeat of the mtDNA was 24,473 bp, the shortest among the tested mtDNAs.

3.2. Gene arrangement and content of Ag0067 mtDNA

Type I and type II of Ag0067 mtDNA had the same content of genes. Regions ID1-ID2-ID3 in both genotypes possessed full sets of typical fungal genes, including three cytochrome *c* oxidase subunits (*cox*1, *cox*2, and *cox*3), three ATP synthase subunits (*atp*6, *atp*8, and *atp*9), seven NADH dehydrogenase subunits (*nad*1, *nad*2, *nad*3, *nad*4, *nad*4L, *nad*5, and *nad*6), a cytochrome *b* (*cob*), and a *rps* gene. It encoded ribosomal RNA of large and small subunits (rns and rnl), and 27 tRNAs. These tRNA genes coded for all 20 common amino acids, including three copies for each of *trn*L and *trn*M, two copies for each of *trn*F, *trn*R and *trn*S, and a single copy for each of other amino acids. Two clusters of tRNA genes were identified, which located in the upstream of the *rns* and *rps3* genes, respectively. For type I, genes located in ID2 and in the 60,989–67,395 bp segment of ID3 were in the same direction, whereas the orientation of the other genes was opposite (Fig. 2A). Compared to type I mtDNA, the orientation of ID1 genes in type II was opposite. ID4 was an identical inverted repeat of ID2, which contained four protein-coding genes (*atp*8, *nad*1, *nad*6, and *rps*3), as well as genes encoding *rnl* and nine tRNAs (*trn*A, *trn*E, *trn*G, *trn*M, *trn*N, *trn*P, *trn*S, *trn*V, and *trn*Y). Therefore, the 14 genes were duplicated in mtDNA of the Ag0067 isolate.

Similar to the mtDNA of Ag0067, the ID1-ID2-ID3 and ID4 for the other isolates sequenced in this study connected end-to-end, with identical flanking sequences of junctions. The ID1-ID2-ID3 of the isolates contained the same sets of protein-coding genes and rRNA genes. The number of tRNA genes possessed by the *A. aegerita* isolates ranged from 25 to 30. The ID4 sequences of all the *A. aegerita* isolates were reverse complementary to their corresponding ID2 sequences. Compared with the other isolates, the ID4 in the SWS17 isolate lost a fragment between *trn*V and *rps*3, including five genes namely *rps*3, *trn*A, *trn*E, *trn*P, and *trn*V. As a result, the ID4 in the SWS17 isolate just possessed nine core genes. The ID1-ID2-ID3 sequences for all the nine isolates contained all information of their mtDNAs, and were taken into account in the following analyses.

3.3. Mitochondrial genes of linear-plasmid origin

The mitochondrial genome of Ag0067 possessed two B-type DNA polymerase genes (*polB* A and *polB* B) and two small orthologs (SO A and SO B) with linear-plasmid origin (Fig. 3A). *polB* A gene was a 1944-bp ORF located between positions 53,623 and 55,566. *polB* B located between positions 49,944 and 51,986, was a disrupted gene due to shift mutation. SO A was a 771-bp ORF (ORF256) located between positions 14,463 and 15,233, while SO B located between positions 10,741 and 10,833. Sequence alignment with plasmid-like fragment in WT-11 mtDNA showed that Ag0067 contained one more 35,872-bp fragment (IS) located 100-bp downstream of *polB* B gene (Fig. 3A and B). The IS included ID2 region, as well as *cox2* and two tRNA genes.

Except for Ag0058, SWS17 and Ag0002, all the other isolates also contained two copies of *polB* genes, which were corresponding to *polB* A and *polB* B of Ag0067 based on their locations (Supplementary Table 2). Only one *polB* gene was detected in isolates Ag0058, SWS17, and Ag0002. Phylogenetic analysis demonstrated that all *polB* genes of Ag0058, SWS17, and Ag0002 were clustered into the group of *polB* A. *polB* p1 of WT-11 was closer to proteins in the group of *polB* B. All the other isolates contained two copies of SO genes, which were corresponding to SO A and SO B of Ag0067 based on their locations (Supplementary Table 2). All SO genes were clustered into two groups, clade SO A and SO B. Small ortholog of WT-11 was clustered into clade SO A.

3.4. Intra-specific diversity of mitochondrial core genes

Intra-specific polymorphisms of each conserved protein-coding gene (CPG) were assessed respectively (Supplementary Table 3). The CDS lengths of most of the CPGs were identical, except for *nad3*, *nad4*, *nad5*, and *rps3*. The length polymorphisms of the four genes presented in their 3′-ends, leading to mutation of stop codon. The density of SNP in all CPGs ranged from 0 to 17.77 SNPs/Kb. The genes *atp8* contained no mutation, *nad6* contained a non-synonymous, and *atp9* contained a synonymous mutation. These represented the lowest nucleotide diversities of the coding sequences (π = 0, π = 0.0006, and π = 0.001). Non-synonymous variations were observed in all the other genes, with the highest

Table 1
Mitochondrial genomic information of the eight *A. aegerita* isolates sequenced in this study together with the publicly available SWS17. ID1-ID2-ID3 and ID4 linked end-to-end, forming a circular mtDNA. *, gene length excluded introns.

Regions	Isolates	Total length	Gene		Intron		Intergenic region
			Length*	Number	Length	Number	
(ID1-ID2-ID3)	Ag0002	92,992	24,716	47	11,470	12	56,806
	Ag0033	98,890	24,562	44	22,791	21	51,537
	Ag0058	99,773	24,584	45	27,003	23	48,186
	Ag0065	102,479	24,603	45	22,628	16	55,248
	Ag0067	90,311	24,649	43	9701	10	55,961
	Ag0072	103,439	24,478	43	16,587	16	62,374
	Ag0085	96,624	24,485	43	12,479	14	59,660
	Ag0086	102,575	24,485	43	16,450	16	61,640
	SWS17	91,856	24,317	42	7458	9	60,081
ID4	Ag0002	30,409	10,483	14	6380	5	13,531
	Ag0033	35,145	10,517	14	7837	9	16,791
	Ag0058	34,918	10,516	14	8121	9	12,536
	Ag0065	30,440	10,500	14	6172	7	13,768
	Ag0067	31,173	10,516	14	6847	7	17,555
	Ag0072	30,440	10,500	14	6172	7	13,768
	Ag0085	28,899	10,505	14	4958	6	13,436
	Ag0086	28,968	10,506	14	4832	6	13,629
	SWS17	24,473	8657	9	4482	5	11,334

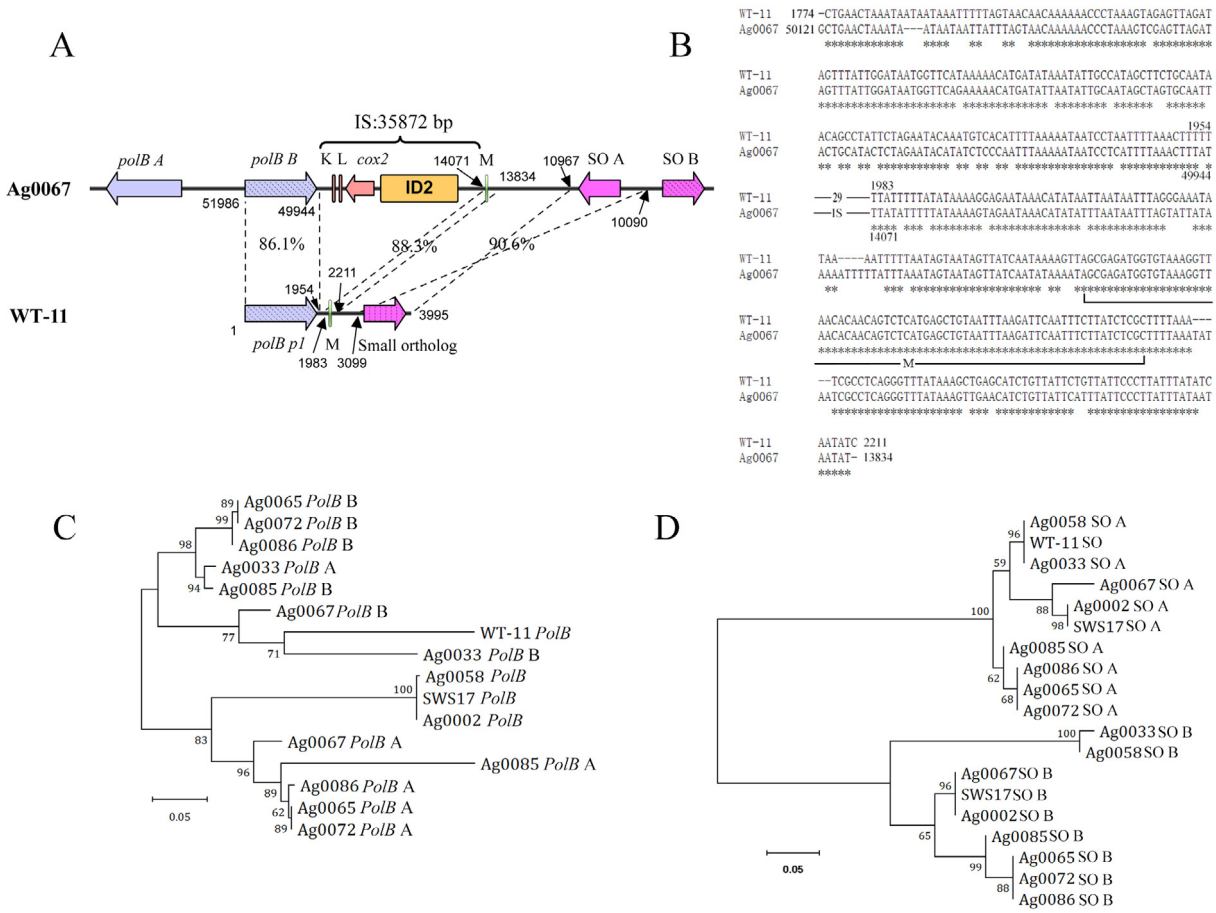


Fig. 3. Plasmid-origin sequences in *A. aegerita* mtDNAs. *polB* A, *polB* B, and *pol*B p1, B-type DNA polymerase genes. SO, SO A and SO B, small orthologs with plasmid origin. IS, the 35,872 bp fragment containing ID2 region, *cox2* gene, *trnK* and *trnL* genes. K, L and M represent tRNA genes of *trnK*, *trnL*, *trnM*. A) Collinearity of plasmid-origin sequences in mtDNAs of isolates Ag0067 and WT-11 (H4 region). B) Alignment of sequences flanked by IS of between Ag0067 and WT-11. C) Phylogenetic tree constructed with amino acid sequences of B-type DNA polymerase genes. D) Phylogenetic tree constructed with amino acid sequences of small orthologs.

non-synonymous/ synonymous polymorphic rate in *nad3* and *nad4L* (Supplementary Table 3).

A phylogenetic tree was constructed based on alignments of the concatenated nucleotide sequences of the 14 CPGs (the *rps3* gene excluded), using the sequences from *C. perniciosa* as an outgroup (Fig. 4 left). Based on a total number of 1764 polymorphic sites across 13,141 aligned base pairs, the mitochondrial sequences were clustered into two major clades. All of the three cultivated isolates Ag0002, Ag0033, and SWS17, as well as the two wild isolates Ag0058 and Ag0067 collected from Jianning, formed a clade (Clade I). The other clade (Clade II) consisted of one wild isolate from Jianning, and three wild isolates collected from other locations.

Fifty nine polymorphic sites were detected in 6420 aligned base pairs of the *rnl* gene, resulting in a SNP density of 9.19 SNPs/Kb. In addition, 17 indel events with 43-bp gaps or missing data were identified in the alignment of the *rnl* genes from the nine isolates. The SNP density of the *rns* genes from the nine isolates was 10.10 SNPs/Kb. The evolutionary rate of the *rnl* and *rns* genes was similar. It was in the middle level of the evolutionary rate of the CPGs determined in this study.

The mitochondrial genome of the Ag0002 isolate possessed 30 tRNAs, which was the highest among all the test isolates. A total of 25 tRNA genes located in the regions of Ag0002:14001–50000 and Ag0002:61001–92992. The polymorphism of these genes among all the isolates was assessed. Five polymorphic sites were detected in 1876 aligned base pairs, resulting in a SNP density of 2.67 SNPs/Kb. The density was lower than that of rRNA genes and most CPGs. Five tRNA genes located in the other two regions (Ag0002:1–14000 and Ag0002:50001–61000) were variable among different isolates (Fig. 4). None of the tRNA genes were found to cover all of the isolates. Corresponding locations among different isolates may contain different contents of tRNA genes. For example, a *trn*F gene present in the 866–935 bp region of Ag0067, whereas it was *trn*L in Ag0058. Sequence comparison showed that insertion or deletion of DNA fragments can lead to gain or loss of the tRNA genes. The SNP densities of the tRNA genes corresponding to the 863–932, 11574–11647, 50450–50534 and 60116–60187 positions of the Ag0002 isolate were 28.6, 82.2, 200.0, and 13.9 SNPs/Kb, respectively. The SNP densities of the tRNA genes were much higher than those of the CPGs and the rRNA genes.

3.5. Intra-specific diversity of introns

The conserved genes of *A. aegerita* mtDNAs possessed abundant introns. Twenty-seven intron-inserted sites were identified from the nine isolates, including ten from *rnl*, four from *rns*, three from *cob*, two from *nad5*, and one from *nad*1, and one from *cox*2 (Table 2). Based on sequence length and similarity (threshold value = 90%), introns inserted into each of nine sites were separated into two to six types (Table 2). For example, the introns inserted in the 227-bp position of *rnl* ranged from 99 to 303 bp in size, including five types. In total, 47 introns distributed in

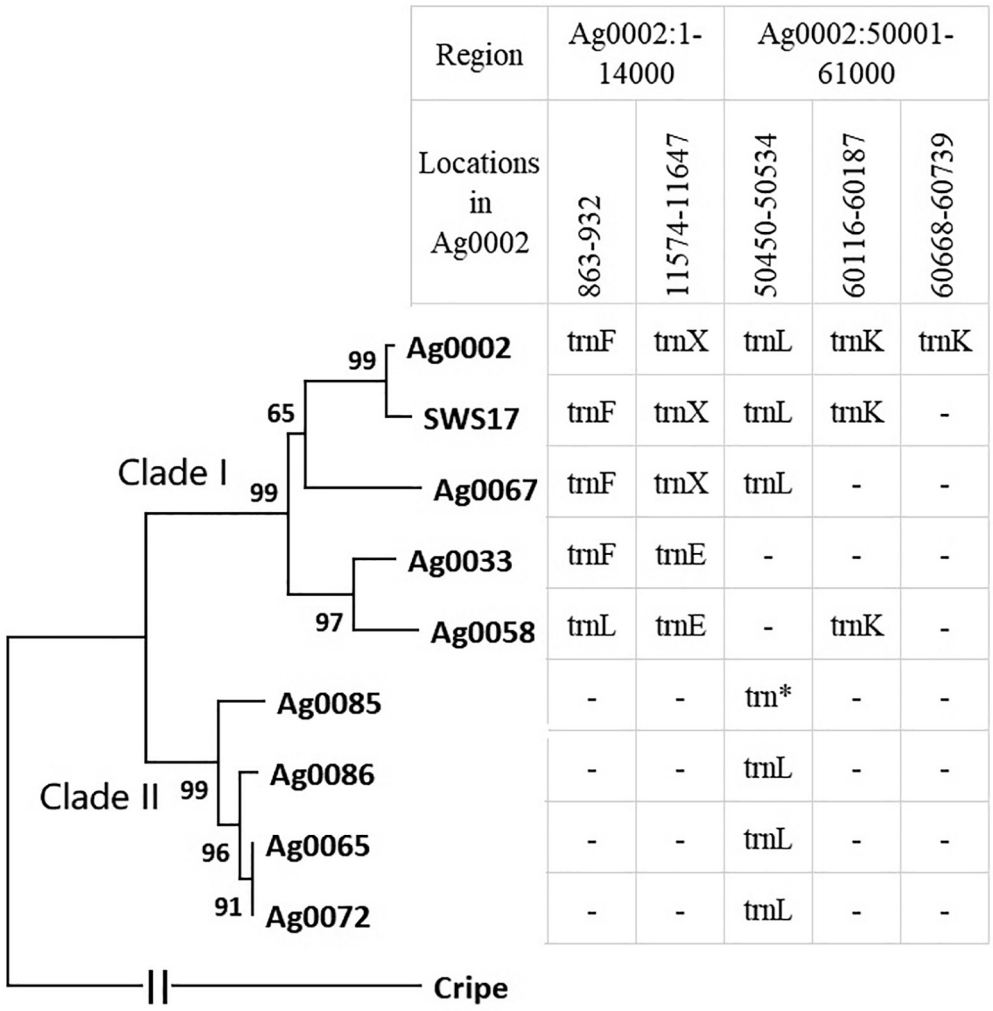


Fig. 4. Intra-specific diversity of tRNAs in the corresponding regions of Ag0002:1–14000 and Ag0002:50000–61000. Left, phylogenetic tree constructed based on 14 CPGs; right, distribution of tRNAs. Clade I and Clade II, two main clades of the phylogenetic tree; -, no tRNA genes; *, unknown types of tRNA.

two rRNA and five protein-coding genes. Twenty-eight out of these introns belonged to the family of group I type. Type II of rns-i2 was a group II intron, but contained an ORF encoding L1 HEG. Five types of rnl-i4 did not contain any conserved RNA secondary structure, but carried a degenerated gene coding for a homolog of double L1. Fourteen small introns were shorter than 303 bp and belonged to undetermined intron types.

The number of introns varied greatly among all the tested isolates, as few as 9 in SWS17 while as many as 25 in Ag0058. There were no introns that were shared between all isolates. Closely related isolates did not always share similar numbers and sequences of introns. For example, the intron number between Ag0067 and Ag0058 differed greatly. The presence/absence of a lot of introns was inconsistent with the evolutionary relationship among the isolates. These features demonstrated a fast movement of introns in *A. aegerita*.

3.6. Intra-specific diversity of gene order

The order of PCGs, rDNA, and common tDNA genes in all the mtDNAs were compared to assess the intra-specific diversity of mtDNA structure (Fig. 5). Two types of gene order were observed. Based on the phylogenetic tree, the nine isolates were divided into two major clades, Clade I and Clade II (Fig. 4 left). Clade I isolates, including Ag0002, Ag0033, Ag0058, SWS17 and Ag0067, shared the same gene order. The orientation of *atp*8, *nad*6, *nad*1, and *cox*2 was counterclockwise, while other genes were arranged clockwise. Compared with the gene order in Clade I isolates, an inversion took place in approximate 10-Kb region between *cox*2 and *nad*4L gene of mtDNAs in Clade II isolates (Ag0065, Ag0072, Ag0085, and Ag0086). This region contained one rRNA gene and a cluster of tRNA genes (R-M-S-H-C-D-F-W-R-T).

3.7. Types and amplification of short insertions/deletions.

Multiple alignments of the mtDNAs revealed 64 types of insertion /deletion sequences shorter than 200 bp, namely short insertions/deletions (SI/Ds) (Supplementary Table 4). Among these SI/Ds, 74% were shorter than 100 bp, and the other 26% ranged from 100 bp to 176 bp. Most of the SI/Ds (86%) contained at least one palindrome; five out of the other nine were tandem repeat sequences. Some SI/Ds varied in length, but shared high similarity in sequences. Using 90% sequence similarity as a threshold, the 64 SI/Ds were divided into 41 clusters (Supplementary Table 5). Among them, three clusters had four members for each, three clusters had three members for each, and seven clusters had two members for each. For example, the four members I1, I33, I43, and I46 in cluster 1 were 32, 41, 65, and 157 bp in length, respectively, all of which contained a common 32-bp sequence. Common sequences

Table 2
Intron information of the nine *A. aegerita* isolates. # intron number in the corresponding isolates. * Degenerated conserved domain. Intact ORF, ORF with complete conserved domains.

Intron	Type	Position (aa or nt)	Number and length of introns									Sequence similarity	Intron type	conserved domains	Intact ORF
			Ag0002 13#	SWS17 9#	Ag0067 11#	Ag0033 22#	Ag0058 25#	Ag0085 15#	Ag0086 17#	Ag0065 17#	Ag0072 17#				
rnl-i1	I	681	99	99	–	–	–	–	–	–	–	100	–	–	–
	II		–	–	169	169	–	–	–	–	–	100	–	–	–
	III		–	–	–	–	171	–	–	–	–	–	–	–	–
	IV		–	–	–	–	–	303	–	–	–	–	–	–	–
	V		–	–	–	–	–	–	186	186	186	100	–	–	–
rnl-i2	I	2324	–	–	32	32	31	32	32	32	32	96.9	–	–	–
rnl-i3	I	2329	–	–	34	34	–	–	–	–	–	100	–	–	–
	II		–	–	–	–	173	–	–	–	–	–	–	–	–
rnl-i4	I	2901	–	–	861	–	–	–	–	–	–	–	–	2L1*	–
	II		–	–	–	1072	–	–	–	–	–	–	–	2L1*	–
	III		–	–	–	–	–	–	–	759	759	100	–	2L1*	–
	IV		–	–	–	–	–	–	764	–	–	–	–	2L1*	–
	V		–	–	–	–	–	652	–	–	–	–	–	2L1*	–
	VI		9	9	–	–	9	–	–	–	–	100	–	–	–
rnl-i5	I	3909	2175	2175	2175	2175	–	–	–	–	–	100	IA	2L1	ORF326
	II		–	–	–	–	2233	–	–	–	–	–	IA	2L1*	–
	III		–	–	–	–	–	8	8	8	8	100	–	–	–
rnl-i6	I	4045	–	–	–	–	1113	1120	1113	1113	1113	99.3	I (derived, B1)	L1	ORF189
rnl-i7	I	4257	–	–	1344	1344	1344	–	–	1344	1344	99.8	IC2	2L1	ORF393
rnl-i8	I	4866	1905	–	–	1904	–	–	–	–	–	99.9	IB	2L1	ORF330
	II		–	–	–	–	1945	–	–	–	–	–	IB	2L1	–
	III		–	–	–	–	–	–	1971	1971	1971	99.9	IB	2L1	–
	IV		–	–	–	–	–	2080	–	–	–	–	IB	2L1	–
rnl-i9	I	5204	1475	1475	–	–	–	–	–	–	–	100	IA3	2L1	ORF327
	II		–	–	1503	1503	1503					99.9	IA3	2L1	–
rnl-i10	I	5291	–	–	–	1280	1280	–	–	–	–	100	IC2	2L1	0RF375
nad1-i1	I	120	724	724								100	IB	–	–
	II					798	791					98.4	IB	–	–
	III				770							–	IB	–	–
	IV							802	797	791	791	96.7	IB	–	–
cox2-i1	I	146	–	–	–	1329	1329	–	1329	1329	1329	99.5	IB	L2	–
rns-i1	I	521	1372	–	–	1372	1372	1372	1372	1372	1372	99.5	IC2	2L1	ORF424
rns-i2	I	1041	35	–	–	35	–	35	–	–	–	100	–	–	–
	II		–	1575	1579	–	1584	–	1574	1575	1575	99.1	II (domainV)	L1	ORF256
rns-i3	I	1291	89	89	–	–	–	–	–	–	–	100	–	–	–
rns-i4	I	1794	44	44	44	44	44	–	–	–	–	100	–	–	–
			–	–	–	–	–	84	84	84	84	100			
nad5-i1	I	239	–	–	–	1208	1207	–	1204	–	–	99.5	IB	L2	ORF274
nad5-i2	I	318	–	–	1249	–	–	–	–	–	–	–	IB	2L1*	–
cox1-i1	I	126	–	–	–	1322	1322	–	–	–	–	99.9	IB	L2	ORF393
cox1-i2	I	242	–	–	–	1036	1036	1036	1036	1036	1036	99.8	IB	2L1	ORF310
cox1-i3	I	322	–	–	–	–	1261	–	–	–	–	–	IB (extra group)	L2	ORF275
cox1-i4	I	351	1268	1268	–	1268	1268	1268	–	–	–	99.8	IB	grplintron_endo	ORF391
cox1-i5	I	367	–	–	–	1008	1008	1009	–	–	–	99.6	I (derived, B2)	2L1	ORF220
cox1-i6	I	433	1152	–	–	1152	1152	–	1152	1152	1152	99.4	IA	grplintron_endo	ORF269
cob-i1	I	67	–	–	–	1614	1614	1604	1606	1607	1607	96.4	IB	L1	ORF193
cob-i2	I	169	1148	–	–	1148	1148	–	1148	1148	1148	99.7	IB	2L1	–
cob-i3	I	274	–	–	–	–	1129	1129	1129	1129	1129	99.6	IB	L2	–

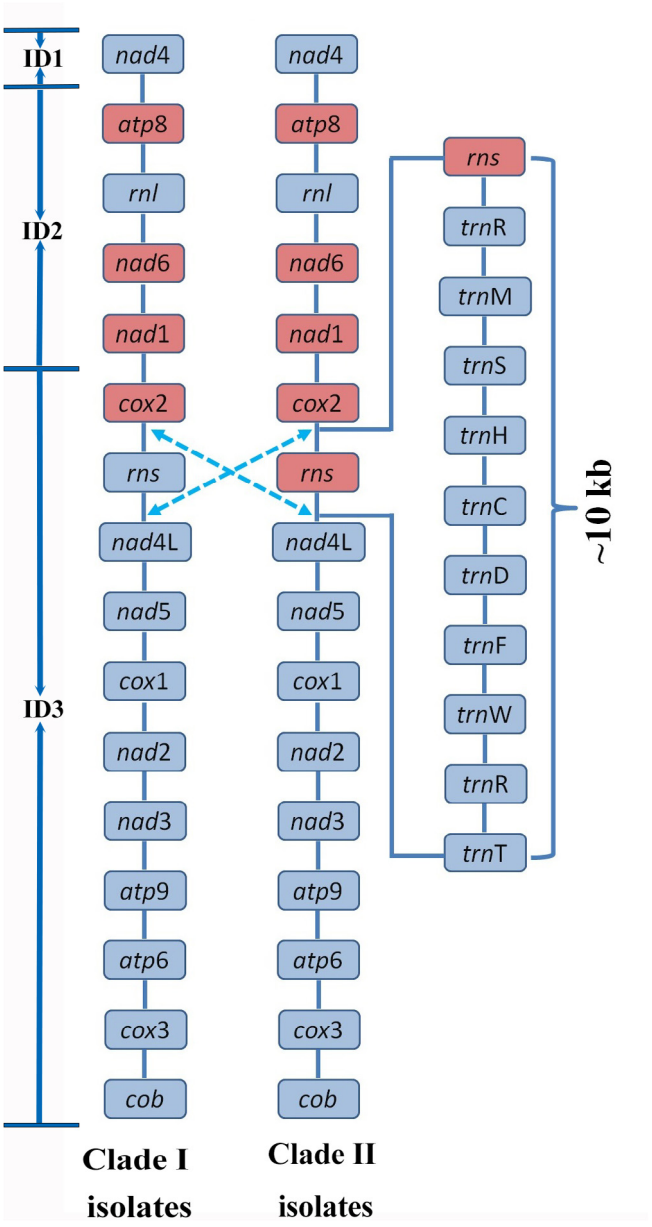


Fig. 5. Gene order of the nine mitochondrial genomes. Clade-I isolates and Clade-II isolates, gene orders for isolates in corresponding clades shown in Fig. 4. Light-blue boxes, clockwise genes; brown boxes, counter-clockwise genes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

for all clusters were palindrome containing or tandem repeat sequences (Supplementary Table 5).

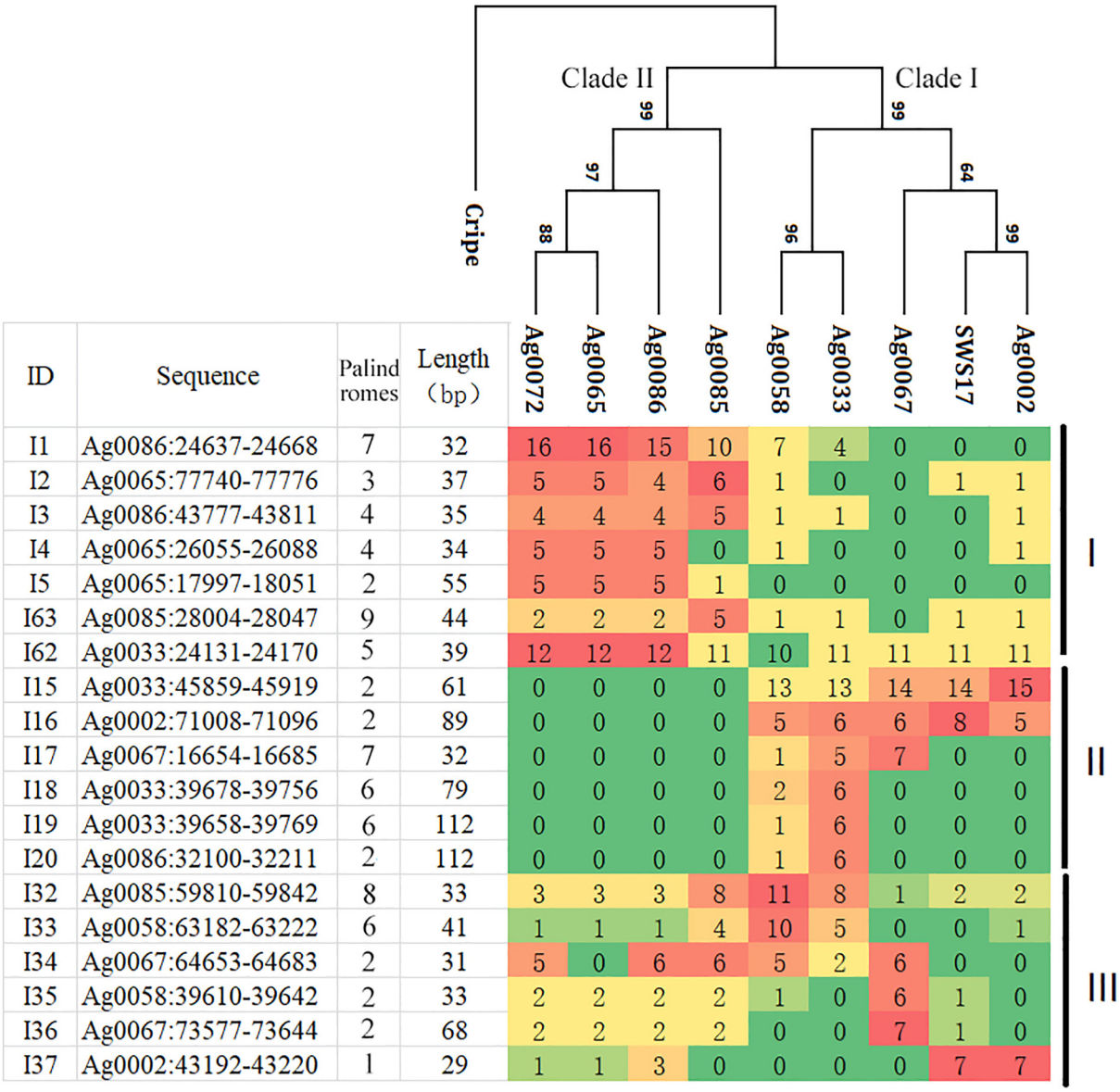
Copy number of some SI/Ds varied greatly among the isolates (Fig. 6). Nineteen SI/Ds had at least five copies in the mtDNAs of one or more isolates. Seven SI/Ds (I in Fig. 6) presented in the clade-II isolates had higher copy numbers than the clade-I isolates, while six SI/Ds (II in Fig. 6) were not detected in mtDNAs of the clade-II isolates. Distribution of another six SI/Ds (III in Fig. 6) was not completely consistent with the evolutionary relationship of the isolates. A noticeable feature of the SI/Ds was a rapid proliferation within genome after a short insertion invasion. The I1 and I15 SI/Ds were two typical examples for this case. High copies of I1 were identified in the mtDNAs of Ag0065, Ag0072, and Ag0086, whereas none were identified in Ag0002 and Ag0067. Similarly, the mtDNAs of the clade-II isolates had no I15 SI/D, whereas the clade-I isolates carried 13 copies or more. SIRs presented in all of the short inserts.

Introns of the same inserted site varied in length, but shared high similarity in sequences. It resulted mainly from the movement of IS/Ds. The introns from nine insertion sites were separated into multiple types according to their lengths and sequence similarities. Most of the insertion sites were located in the rRNA genes. Five types of introns were identified in a site of the 227-bp position of the *rnl* gene (rnl-i1, Fig. 7A). The I34 and I62 SI/D**s** invaded into the sites of the 10-bp and 26-bp positions in the Type I, respectively, forming a second type of the intron. The 42-bp sequence in the 5′-end of the Type II was replaced by the I63, forming a Type III. Compared with the Type V, the Type IV contained an additional 119-bp fragment at the 3′-end, which was a complex short intron of I64 being embedded by I63. Rnl-i4 also had five intron types (Fig. 7B). Compared with the Type I, the Type II intron contained three repeats of I18, which was inserted into the sites of the 42-bp, 129-bp, and 236-bp positions in the Type I, respectively. An I65 was inserted into the 100-bp site of the Type V, forming a Type IV intron. Intron Rnl-i5 in the Ag0002, Ag0033, Ag0067, and SWS17, with an identical sequence of 2175 bp length, all belonged to the group IA intron. It contained an ORF326 encoding a double L1 HE (Fig. 7C). An I47 inserted into a 911-bp site, forming a rnl-i5 intron of the Ag0058. Consequently, the double L1 HE gene was degenerated. Besides, insertions of one or more short fragments led to length polymorphism of the introns from rnl-i3, rnl-i8, rnl-i9, nad1-i1, and rns-i2.

4. Discussion

In this study, two genotypes of mtDNAs in Ag0067 isolate were determined by mapping of long PacBio reads. Both genotypes are in a quadripartite structure consisting of a pair of identical inverted repeats separated by two single-copy regions. The only difference between them was the orientation of the single-copy regions. The type of rearrangement is analogous to chloroplast genomes, in which two inverted repeats are divided by small and large single-copy regions [50]. Chloroplast genomes in most plants also possess two structural haplotypes differing in the orientation of single-copy regions [51]. Sequencing depth of illumina reads has been widely used in assessing copy number of certain regions in a genome, such as rDNA repeat units [52]. In this study, the Illumina sequencing depth of the inverted repeats in all the mtDNAs was nearly twice that of the other regions, indicating that the region was duplicated in the sequenced genomes. The duplicated segments of all the isolates except that SWS17 located in the same region, containing same sets of protein-coding genes, rRNA genes and common flanking sequences of junctions. The region corresponded to the inverted repeat of Ag0067. These suggest that mtDNAs of these isolates have the same genomic structure as Ag0067. The pairs of inverted repeats range from 57.8 Kb to 70.2 Kb in size, half the size of the complete mitochondrial genomes. mtDNAs containing large inverted repeats have been detected in nine species of *Termitimyces* [53]. The inverted repeat sequences were determined by analyzing sequencing depth of short reads, as well as confirmation of IR border using PCR amplifications [53]. Benefited from the advantages of long PacBio reads, we further constructed the quadripartite architecture of Ag0067 mtDNAs with large inverted repeats.

The mtDNA of *A. aegerita* isolate SWS17 has been sequenced in a previous study using Illumina HiseqX-10 system [35]. The 116,329 bp mtDNA also contained a large inverted repeat of 24,473 bp, which is shorter than all of the eight sequenced mtDNAs from this study. Sequence comparison demonstrated that the inverted repeat in SWS17 lacked a fragment at its 5′-end, which



ID	Sequence	Palindromes	Length (bp)	Ag0072	Ag0065	Ag0086	Ag0085	Ag0058	Ag0033	Ag0067	SWS17	Ag0002	
I1	Ag0086:24637-24668	7	32	16	16	15	10	7	4	0	0	0	
I2	Ag0065:77740-77776	3	37	5	5	4	6	1	0	0	1	1	
I3	Ag0086:43777-43811	4	35	4	4	4	5	1	1	0	0	1	
I4	Ag0065:26055-26088	4	34	5	5	5	0	1	0	0	0	1	I
I5	Ag0065:17997-18051	2	55	5	5	5	1	0	0	0	0	0	
I63	Ag0085:28004-28047	9	44	2	2	2	5	1	1	0	1	1	
I62	Ag0033:24131-24170	5	39	12	12	12	11	10	11	11	11	11	
I15	Ag0033:45859-45919	2	61	0	0	0	0	13	13	14	14	15	
I16	Ag0002:71008-71096	2	89	0	0	0	0	5	6	6	8	5	
I17	Ag0067:16654-16685	7	32	0	0	0	0	1	5	7	0	0	II
I18	Ag0033:39678-39756	6	79	0	0	0	0	2	6	0	0	0	
I19	Ag0033:39658-39769	6	112	0	0	0	0	1	6	0	0	0	
I20	Ag0086:32100-32211	2	112	0	0	0	0	1	6	0	0	0	
I32	Ag0085:59810-59842	8	33	3	3	3	8	11	8	1	2	2	
I33	Ag0058:63182-63222	6	41	1	1	1	4	10	5	0	0	1	
I34	Ag0067:64653-64683	2	31	5	0	6	6	5	2	6	0	0	
I35	Ag0058:39610-39642	2	33	2	2	2	2	1	0	6	1	0	III
I36	Ag0067:73577-73644	2	68	2	2	2	2	0	0	7	1	0	
I37	Ag0002:43192-43220	1	29	1	1	3	0	0	0	0	7	7	

Fig. 6. Distribution of SI/Ds with high copy number. ID numbers in the first column are corresponding to those of Supplementary Table 4. Clade I and Clade II, two main clades of the phylogenetic tree. Number in column 3, number of palindromes carried by the corresponding SI/D. Colored numbers represent copy number of SI/Ds in corresponding mtDNAs. Color gradual from green to dark red, copy number of SI/Ds from zero to high. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

contained five core genes. Short Illumina reads are generally difficult to assemble as long repeat sequences in the genome. The missing fragment is not sure to be duplicated due to lacking of sequencing-depth data. The phylogenetic tree showed that the relationship between SWS17 and Ag0002 was closer than that between Ag0002 and the other tested isolates, indicating that SWS17 mtDNA was similar to Ag0002 in gene content and organization. It suggested that the inverted repeat in the SWS17 isolate contained the same set of core genes with the other tested isolates, while the lack of the fragment possibly resulted from wrong assembly.

Ferandon, Chatel et al. [33] identified two large inverted repeats separated by a region containing plasmid-origin genes (H4) in the mtDNA of an *A. aegerita* isolate WT-11. Each of the repeats contained identical copies of *nad4* gene. In comparison, the ID4 of all the isolates reported in our study had conserved gene content, including four protein-coding genes *atp*8, *nad*1, *nad*6, and *rps*3, but no *nad4*. The Illumina sequencing depth of ID1-ID2-ID3 revealed that none of the tested isolates contained repeats of *nad4* gene. Sequence similarity of *nad4* as compared between WT-11 and each of the tested isolates (<96%) were lower than compared among the tested isolates in this study (>99%). These results indicated that the mtDNA of WT-11 had a different genomic structure than the tested isolates. The phylogenetic analysis based on sequences of mtSSU-rDNA V4-V6-V9 domains revealed three clades within the *A. aegerita* complex: a European, an Argentinean, and an Asian-American clade [31]. All the tested isolates belonged to the Asian-American clade, whereas WT-11 clustered with the European clade. The genomic structure variation is supposed to occur after divergence between the European and the Asian-American clades.

The H4 region of WT-11 mtDNA contained a *polB* gene and a small ortholog, and a *trnM* gene between them [33]. Both polB gene and small ortholog originated from the same plasmid. Collinearity

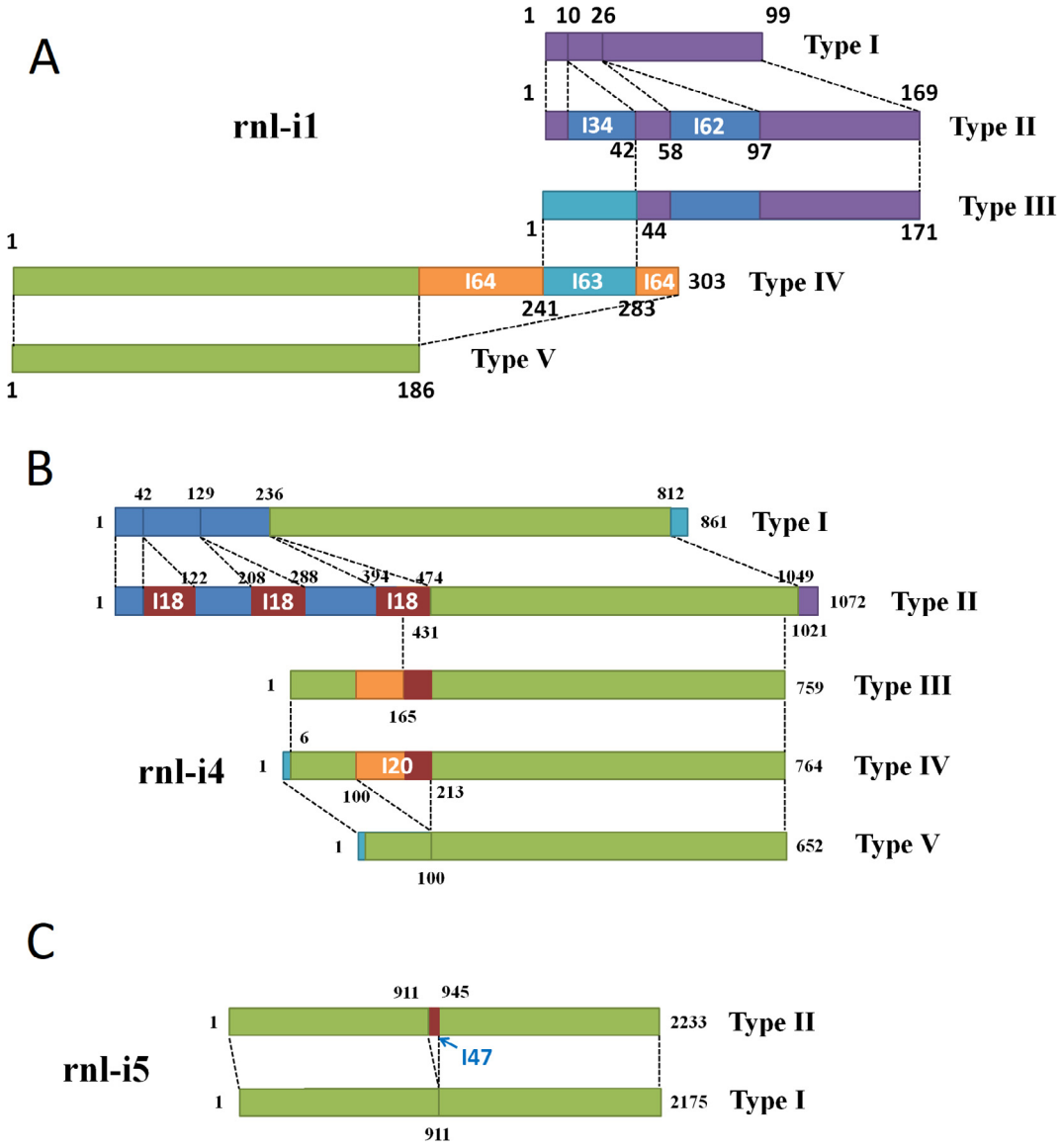


Fig. 7. Comparison of intron types within the same insertion sites of rnl-i1 (A), rnl-i4 (B), and rnl-i5 (C). Different insertions were represented by different colors. The numbers in the diagrams represent positions of introns.

analysis as well as sequence alignment with H4 revealed that there was an extract large segment (IS) located between *polB* and *trnM* genes of Ag0067 mtDNA. IS contained whole sequence of ID2. Without the whole mitochondrial sequence of WT-11 isolate, it is uncertain if the IS resulted from a large insertion or inversions. Whatever the source, at least one event occurred in the region after divergence of WT-11 and Ag0067.

This study provides a hint for putative intramolecular recombination events within mtDNAs. PacBio sequencing revealed that Ag0067 contained two mtDNA genotypes. Given that the mycelium used for sequencing in this study was originated from the isolate of a single hyphae tip, it suggests that one Ag0067 compartment harbored both genotypes of mtDNAs. Duo to the highly similar architectures, a hypothesis, known as flip-flop recombination to explain the presence of structural heteroplasmy of chloroplast genomes [54,55], may also apply to Ag0067 mtDNAs. In the hypothesis, large inverted repeats could mediate intramolecular recombination within one genotype, which lead to formation of the other genotype (Fig. 8). Given the identical sequences of the inverted repeats, the intramolecular recombination events always took place before the mutagenesis of both repeat sequences. In combination with the phenomenon of two genotypes coexisting in the same compartment, this indicates that the recombination was frequent. The frequency of type I to type II genotype was more than two, which was different from the equal frequency between two chloroplast haplotypes in most plant individuals [51]. A possible explanation is that few reads were able to span in the IR region larger than 31 kb, which were used to distinguish genotypes.

Non-mitochondrial tRNAs are commonly regarded as the most conservative genes, while mitochondrial tRNAs (mt-tRNAs) in human and animal are often structurally diverged. It is unclear about the evolutional rate of fungal mt-tRNA, due to few reports in relevant fields. In this study, the tRNA genes in *A. aegerita* mtDNAs were found to undergo two divergent evolutional rates based on the regions they located in. The tRNAs in ID2 and the latter part of ID3 were highly conservative in numbers and sequences, and evolved in a way similar to PCGs and rDNAs, whereas those in the ID1 and the first 11 Kb of ID3 were highly variable in numbers, types, and sequences. Insertions or deletions of the fragments led to gain or loss of tRNAs. All linear plasmid elements located in

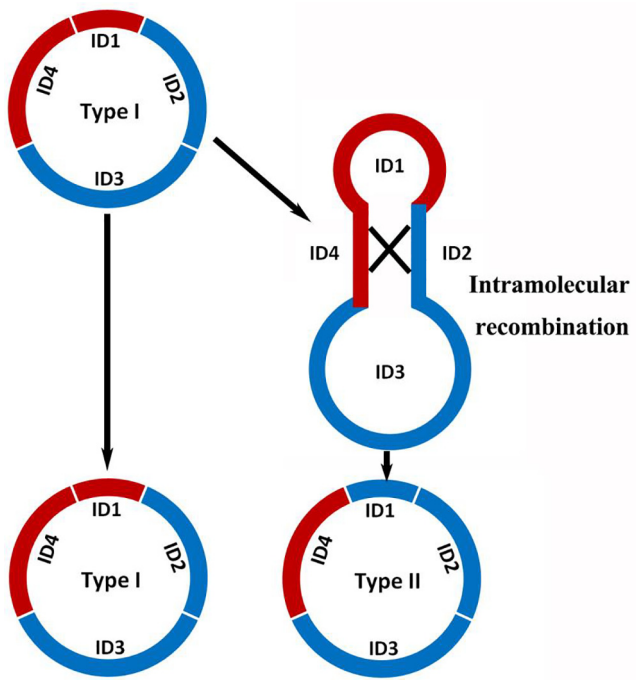


Fig. 8. Hypotheses of flip-flop recombination for formation the other genotype of *A. aegerita* mtDNAs.

the regions of the ID1 and the first 11 Kb of ID3; gene order exchange took place in the first part of ID3. These phenomena indicate that they were highly variable regions. It suggests that fast structural variation was a main force driving the diversity of tRNAs in intraspecific mtDNAs.

Intron variation is a main factor contributing to intra-specific diversity of *A. aegerita* mtDNAs. The variability of mitochondrial intron content between different isolates has been estimated in many fungal species, as mentioned in the introduction. Similarly in *A. aegerita*, Ag0085 carried more than twice the number of introns than the closely related isolate Ag0067. The size and sequences of introns from the same insertion site vary greatly among the isolates. Using a similarity threshold of 90%, six intron types were identified in rnl-i4, five types in rnl-i1, four types in each of nad1-i1 and rnl-i8. Overall, no intron type was presented in all the isolates.

Besides introns, invasion of short insertions is another way to diversify the mitochondrial genomes among *A. aegerita* isolates. Invasion of short insertions produced length-polymorphism of introns, leading to high diversity in six out of the eight introns. Among the 27 intron-inserted sites, 14 located in rDNA genes, the other 13 sites were in PCGs. Eight out of nine polymorphic sites were in rDNA genes, indicating that short insertions tend to invade introns of rDNA genes. An invasion of I47 into rnl-i5 interrupted the structure of double L1 domain, leading to loss of its function. Acquisition of short insertion might affect or even hinder the movement of introns, further enhancing the genomic diversity. High diversity of copy number of short insertions among the isolates indicates their fast proliferation after invasion. The higher number of I18s of the type II in rnl-i4 than the type I is a typical example of fast proliferation of short insertions. The presence of short insertions, as well as its potential effect on mitochondrial genome evolution, might provide new molecular markers for population-level studies.

Palindromes are hairpin elements, which can form a local DNA structure under appropriate conditions namely cruciform [56]. Due to the special structures, palindromes are regarded as transposable elements which could amplify through RNA intermediate and recombination [57]. In this study, palindromes presented with over five copies in all the short insertions, and the insertions in the same cluster shared the same hairpin elements. This suggests that palindromes were in charge of the movement of their host short insertions. The presence of hairpin elements might be related to some functions including mRNA processing, translation, stabilization [58], and DNA recombination [59].

5. Conclusion

Mitochondrial genomes of the tested *A. aegerita* isolates had a quadripartite structure consisting of two large identical inverted repeats, separated by two single copy regions. Due to the presence of the inverted repeats, intramolecular recombination took place in one genotype of mtDNA, formatting a second genotype. As a result, divergent mitochondrial genotypes coexisted in the same fungal cells. The number and distribution of the introns varied greatly among mtDNAs. Fast invasion of short insertions is considered to driving the diversity of introns as well as the other non-coding regions, leading to variation of the mtDNAs. Our analyses also revealed that tRNA diversity among the isolates mainly resulted from structural variation in highly variable regions.

Table 3
Accession numbers of sequencing data of the eight *A. aegerita* isolates and the location of IDs.

Isolates	Next-generation sequencing accession number	Third-generation sequencing accession number	Mitogenome accession number	Mitogenome location			
				ID1	ID2	ID3	ID4
Ag0002	SRR12473746	–	MT364880	1–15286	15287–45695	45696–92992	92993–123386
Ag0033	SRR12474001	–	MT364881	1–11434	11435–46579	46580–98890	98891–134035
Ag0058	SRR12474014	–	MT364882	1–11454	11455–46372	46373–99773	99774–134691
Ag0065	SRR12474015	–	MT364883	1–17287	17288–47728	17289–102479	102480–132919
Ag0067	SRR12474046	SRR12474063	MT364879	1–15068	15069–46241	46242–90311	90312–121484
Ag0072	SRR12474049	–	MT364884	1–18171	18172–48611	48612–103439	103440–133879
Ag0085	SRR12474050	–	MT364885	1–15722	15723–44621	44622–96624	96625–125523
Ag0086	SRR12474052	–	MT364886	1–17461	17462–46429	46430–102575	102576–131542

Availability of data: Raw data of the whole genome sequencing, as well as the mitogenome sequences of *A. aegerita* isolates, were submitted to GenBank under the accession numbers listed in Table 3. ID locations were shown in the table.

CRediT authorship contribution statement

Xinrui Liu: Investigation, Formal analysis, Validation, Visualization, Writing - original draft. **Xiaoping Wu:** Methodology, Resources, Validation. **Hao Tan:** Conceptualization, Validation, Formal analysis, Data curation, Writing - original draft. **Baogui Xie:** Supervision, Resources, Funding acquisition. **Youjin Deng:** Conceptualization, Project administration, Methodology, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the Fujian Edible Fungi Engineering Technology Research Center and the National Fungi Breeding Center (Fujian Division) for providing the experimental facilities. This work was supported by grants from the Natural Science Foundation of Fujian Province (Grant No. 2016J01702), and the Natural Science Foundation of China (Grant Nos. 31670021, 31901119). We are grateful to Professor Meiying Guo of Sanming Fungi Research Institute for donating us the Ag0002 and Ag0033 isolates.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2020.08.022.

References

[1] Burger G, Gray MW, Lang BF. Mitochondrial genomes: anything goes. Trends in Genetics Tig 2003;19(12):709–16.
[2] Sandor S, Zhang Y, Xu J. Fungal mitochondrial genomes and genetic polymorphisms. Appl Microbiol Biotechnol 2018;102(22):9433–48.
[3] Johnston IG, Williams BP. Evolutionary inference across eukaryotes identifies specific pressures favoring mitochondrial gene retention. Cell Systems 2016;2(2):101–11.
[4] James TY, Pelin A, Bonen L, Ahrendt S, Sain D, Corradi N, et al. Shared signatures of parasitism and phylogenomics unite Cryptomycota and microsporidia. Curr Biol 2013;23(16):1548–53.
[5] Losada L, Pakala SB, Fedorova ND, Joardar V, Shabalina SA, Hostetler J, et al. Mobile elements and mitochondrial genome expansion in the soil fungus and potato pathogen Rhizoctonia solani AG-3. FEMS Microbiol Lett 2014;352(2):165–73.
[6] Beaudet D, Nadimi M, Iffis B, Hijri M. Rapid mitochondrial genome evolution through invasion of mobile elements in two closely related species of arbuscular mycorrhizal fungi. PLoS One 2013;8(4):e60768.
[7] Himmelstrand K, Olson Å, Durling MB, Karlsson M, Stenlid J. Intronic and plasmid-derived regions contribute to the large mitochondrial genome sizes of Agaricomycetes. Curr Genet 2014;60(4):303–13.
[8] Kanzi AM, Wingfield BD, Steenkamp ET, Naidoo S, van der Merwe NA. Intron derived size polymorphism in the mitochondrial genomes of closely related Chrysoporthe species. PLoS One 2016;11(6):e0156104.
[9] Lang BF, Laforest M-J, Burger G. Mitochondrial introns: a critical view. Trends Genet 2007;23(3):119–25.
[10] Fan WW, Zhang S, Zhang YJ. The complete mitochondrial genome of the Chanhua fungus Isaria cicadae: a tale of intron evolution in Cordycipitaceae. Environ Microbiol 2019;21(2):864–79.
[11] Busse I, Preisfeld A. Discovery of a group I intron in the SSU rDNA of Ploeotia costata (Euglenozoa). Protist 2003;154(1):57–69.
[12] Klassen R, Meinhardt F. Linear protein-primed replicating plasmids in eukaryotic microbes. Microbial Linear Plasmids 2007;7:187–226.
[13] Férandon C, Xu J, Barroso G. The 135 kbp mitochondrial genome of Agaricus bisporus is the largest known eukaryotic reservoir of group I introns and plasmid-related sequences. Fungal Genet Biol 2013;55:85–91.
[14] Formighieri EF, Tiburcio RA, Armas ED, Medrano FJ, Shimo H, Carels N, et al. The mitochondrial genome of the phytopathogenic basidiomycete Moniliophthora perniciosa is 109 kb in size and contains a stable integrated plasmid. Mycol Res 2008;112(10):1136–52.
[15] Ryohsuke N, Kikuo S, Shin-Ichi K, et al. Basidiomycetous fungus Flammulina velutipes harbors two linear mitochondrial plasmids encoding DNA and RNA polymerases. FEMS Microbiol Lett 2000;190(1):99–102.
[16] Yui Y, Katayose Y, Shishido K. Two linear plasmid-like DNA elements simultaneously maintained in Pleurotus ostreatus. Biochimica et Biophysica Acta (BBA)-Gene Struct Expression 1988;951(1):53–60.
[17] Bois F, Barroso G, Gonzalez P, Labarere J. Molecular cloning, sequence and expression of Aa-polB, a mitochondrial gene encoding a family B DNA polymerase from the edible basidiomycete Agrocybe aegerita. Mole General Genet MGG 1999;261(3):508–13.
[18] Mouhamadou B, Barroso G, Labarere J. Molecular evolution of a mitochondrial polB gene, encoding a family B DNA polymerase, towards the elimination from Agrocybe mitochondrial genomes. Mol Genet Genom 2004;272(3):257–63.
[19] Wang Y, Zeng F, Hon CC, Zhang Y, Leung FCC. The mitochondrial genome of the Basidiomycete fungus Pleurotus ostreatus (oyster mushroom). FEMS Microbiol Lett 2008;280(1):34–41.
[20] Lavrov DV, Maikova O, Pett W, Belikov SI. Small inverted repeats drive mitochondrial genome evolution in Lake Baikal sponges. Gene 2012;505(1):91–9.
[21] Pearson CE, Zorbas H, Price GB, Zannishadjopoulos M. Inverted repeats, stem-loops, and cruciforms: Significance for initiation of DNA replication. J Cell Biochem 1996;63(1):1–22.
[22] Brazda V, Laister RC, Jagelska EB, Arrowsmith CH. Cruciform structures are a common DNA feature important for regulating biological processes. BMC Mol Biol 2011;12(1):33.
[23] Cechova J, Lýsek J, Bartas M, Brazda V. Complex analyses of inverted repeats in mitochondrial genomes revealed their importance and variability. Bioinformatics 2018;34(7):1081–5.
[24] Zhang YJ, Zhang S, Zhang GZ, Liu XZ, Wang CS, Xu JP. Comparison of mitochondrial genomes provides insights into intron dynamics and evolution in the caterpillar fungus Cordyceps militaris. Fungal Genet Biol 2015;77:95–107.
[25] Tao YT, Suo F, Tusso S, Wang YK, Huang S, Wolf JBW, et al. Intraspecific diversity of fission yeast mitochondrial genomes. Genome Biol Evol 2019;11(8):2312–29.
[26] de Queiroz CB, Santana MF, Vidigal PMP, de Queiroz MV. Comparative analysis of the mitochondrial genome of the fungus Colletotrichum lindemuthianum, the causal agent of anthracnose in common beans. Appl Microbiol Biotechnol 2018;102(6):2763–78.
[27] Diyabalanage T, Mulabagal V, Mills G, DeWitt DL, Nair MG. Health-beneficial qualities of the edible mushroom Agrocybe aegerita. Food Chem 2008;108(1):97–102.
[28] Liu W, Yu G, Yu W, Ye X, Jin Y, Shrestha A, et al. Autophagy inhibits apoptosis induced by agrocybe aegerita lectin in hepatocellular carcinoma. Curr Med Chem 2017;17(2):221–9.
[29] Jing H, Zhang Q, Liu M, Zhang J, Zhang C, Li S, et al. Polysaccharides with antioxidative and antiaging activities from enzymatic-extractable mycelium by agrocybe aegerita (Brig.) Sing. Evid-Based Complem Alternative Med 2018:1–11.
[30] Motoshima RA, Rosa TF, Mendes LC, da Silva EV, Viana SR, do Amaral BS, et al. Inhibition of Leishmania amazonensis arginase by fucogalactan isolated from Agrocybe aegerita mushroom. Carbohydr Polym 2018;201:532–8.
[31] Uhart M, Sirand-Pugnet P, Labarere J. Evolution of mitochondrial SSU-rDNA variable domain sequences and rRNA secondary structures, and phylogeny of the Agrocybe aegerita multispecies complex. Res Microbiol 2007;158(3):203–12.
[32] Moulinier T, Barroso G, Labarère J. The mitochondrial genome of the basidiomycete Agrocybe aegerita: molecular cloning, physical mapping and gene location. Curr Genet 1992;21(6):499–505.
[33] Ferandon C, Chatel SEK, Castandet B, Castroviejo M, Barroso G. The Agrocybe aegerita mitochondrial genome contains two inverted repeats of the nad4 gene arisen by duplication on both sides of a linear plasmid integration site. Fungal Genet Biol 2008;45(3):292–301.
[34] Barroso G, Blesa S, Labarere J. Wide distribution of mitochondrial genome rearrangements in wild strains of the cultivated basidiomycete Agrocybe aegerita. Appl Environ Microbiol 1995;61(4):1187–93.
[35] Xu LM, Hinsinger DD, Jiang GF. The complete mitochondrial genome of the Agrocybe aegerita, an edible mushroom. Mitochondrial DNA Part B 2017;2(2):791–2.
[36] Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 2017;27(5):722–36.
[37] Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 2014;9(11):e112963.
[38] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19(5):455–77.
[39] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Quantitative Biol 2013:1303–3997.
[40] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25(16):2078–9.

[41] Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z, et al. IBS: an illustrator for the presentation and visualization of biological sequences. Bioinformatics 2015;31(20):3359–61.
[42] Yu Y, Ouyang Y, Yao W. shinyCircos: an R/Shiny application for interactive creation of Circos plot. Bioinformatics 2017;34(7):1229–31.
[43] J.D. Thompson T.J. Gibson D.G. Higgins Multiple sequence alignment using ClustalW and ClustalX Current protocols in bioinformatics (1) 2003 2.3. 1–2.3. 22
[44] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 2018;35(6):1547–9.
[45] Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 2009;25(11):1451–2.
[46] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22(22):4673–80.
[47] Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010;26(19):2460–1.
[48] Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006;22(13):1658–9.
[49] Brazda V, Kolomaznik J, Lýsek J, Haronikova L, Coufal J, Stastný J. Palindrome analyser - A new web-based server for predicting and evaluating inverted repeats in nucleotide sequences. Biochem Biophys Res Commun 2016;478(4):1739–45.
[50] Goulding SE, Olmstead RG, Morden CW, Wolfe KH. Ebb and flow of the chloroplast inverted repeat. Mol Genet Genom 1996;252(1):195–206.
[51] Wang W, Lanfear R. Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants. Genome Biol Evol 2019;11(12):3372–81.
[52] Lofgren LA, Uehling JK, Branco S, Bruns TD, Martin F, Kennedy PG. Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. Mol Ecol 2019;28(4):721–30.
[53] Mathijs N, Van PLJJ, Bakker FT, Zwaan BJ, Aanen DK. Enrichment of G4 DNA and a large inverted repeat coincide in the mitochondrial genomes of Termitomyces. Genome Biol Evol 2019;11(7):1857–69.
[54] Stein DB, Palmer JD, Thompson WF. Structural evolution and flip-flop recombination of chloroplast DNA in the fern genus Osmunda. Curr Genet 1986;10(11):835–41.
[55] Bendich AJ. Circular chloroplast chromosomes: the grand illusion. Plant Cell 2004;16(7):1661–6.
[56] Brázda V, Lýsek J, Bartas M, Fojta M. Complex analyses of short inverted repeats in all sequenced chloroplast DNAs. Biomed Res Int 2018:1–10.
[57] Smith DR, Lee RW. The mitochondrial and plastid genomes of Volvox carteri: bloated molecules rich in repetitive DNA. BMC Genom 2009;10(1):132.
[58] Lu S, Wang G, Bacolla A, Zhao J, Spitser S, Vasquez KM. Short inverted repeats are hotspots for genetic instability: relevance to cancer genomes. Cell Reports 2015;10(10):1674–80.
[59] Dieckmann CL, Gandy B. Preferential recombination between GC clusters in yeast mitochondrial DNA. EMBO J 1987;6(13):4197–203.