Computational and Structural Biotechnology Journal 19 (2021) 5987–5999

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

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

Mitogenome-wide comparison and phylogeny reveal group I intron dynamics and intraspecific diversification within the phytopathogen *Corynespora cassiicola*

Qingzhou Ma [a,1], Haiyan Wu [b,1], Yuehua Geng [a], Qiang Li [c], Rui Zang [a], Yashuang Guo [a], Chao Xu [a,*], Meng Zhang [a,*]

[a] Department of Plant Pathology, Henan Agricultural University, Zhengzhou, Henan, China
[b] Analytical Instrument Center, Henan Agricultural University, Zhengzhou, Henan, China
[c] School of Food and Biological Engineering, Chengdu University, Chengdu, China

ARTICLE INFO

Article history:
Received 12 September 2021
Received in revised form 1 November 2021
Accepted 2 November 2021
Available online 03 November 2021

Keywords:
Corynespora cassiicola
Mitochondrial genome
Intraspecific comparison
Group I intron
mtDNA

ABSTRACT

Corynespora cassiicola, the causal agent of an extensive range of plant diseases worldwide, is a momentous fungus with diverse lifestyles and rich in intraspecies variations. In the present study, a total of 56 mitochondrial genomes of *C. cassiicola* were assembled (except two available online) and analyzed, of which 16 mitogenomes were newly sequenced here. All these circular mitochondrial DNA (mtDNA) molecules, ranging from 39,223 bp to 45,786 bp in length, comprised the same set of 13 core protein-coding genes (PCGs), two rRNAs and 27 tRNAs arranged in identical order. Across the above conserved genes, *nad3* had the largest genetic distance between different isolates and was possibly subjected to positive selection pressure. Comparative mitogenomic analysis indicated that seven group I (IB, IC1, and IC2) introns with a length range of 1013–1876 bp were differentially inserted in three core PCGs (*cox1*, *nad1*, and *nad5*), resulting in the varied mitogenome sizes among *C. cassiicola* isolates. In combination with dynamic distribution of the introns, a well-supported mitogenome-wide phylogeny of the 56 *C. cassiicola* isolates revealed eight phylogenetic groups, which only had weak correlations with host range and toxin class. Different groups of isolates exhibited obvious differences in length and GC content of some genes, while a degree of variance in codon usage and tRNA structure was also observed. This research served as the first report on mitogenomic comparisons within *C. cassiicola*, and could provide new insights into its intraspecific microevolution and genetic diversity.

1. Introduction

Corynespora cassiicola (Berk. & M.A. Curtis) C.T. Wei, belonging to the order Pleosporales, class Dothideomycetes, is a ubiquitous ascomycete responsible for target spot epidemics in a wide range of economically important plants, such as cucumber, tomato, soybean, cotton, pomegranate and rubber tree [1–3]. Over recent years, this pathogenic fungus is gradually becoming a major concern in production of these high-value crops, and related reports have also been increasingly frequent around the world [4,5]. Although mostly known as necrotrophic phytopathogens, *C. cassiicola* isolates displayed diverse lifestyles, from saprophyte [6,7] to endophyte [8–10], and even acted as the etiologic agents of some human diseases, e.g., phaeohyphomycosis [11,12].

Like most *Corynespora* spp., the type species *C. cassiicola* is considered strictly asexual due to the lack of observed sexual reproductive structures [1,13]. Nevertheless, plenty of genetic variations were accumulated within this species, and have been largely evaluated by using various DNA fingerprinting methods including ISSR, iPBS and SRAP markers [14–16]. In some cases, the genetic variability of *C. cassiicola* isolates could more or less correlated with their biological features, such as pathogenicity, host specialization or geographical origin, whereas in other cases no correlation was found. A prominent phylogenetic analysis based on the combined data of four loci (ITS, *caa5*, *ga4* and *act1*) revealed that a total of 143*C. cassiicola* isolates from 68 different plant species worldwide were classified into six highly supported clusters, which correlated with origin of host and pathogenicity, but not

* Corresponding authors.
E-mail addresses: chaoxu01@163.com (C. Xu), zm2006@126.com (M. Zhang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2021.11.002

with location of collection [17]. An approximate relationship was observed between toxin classes (based on diversity of cassiicolin isoform) and the phylogenetic clades described by Dixon et al. (2009), except that some strains without a *Cas* gene were distributed over all phylogenetic branches [18]. Moreover, comparative genomic analysis showed that different *C. cassiicola* isolates had varied genome sizes, gene contents or numbers of effectors, but no clear relation was found between these genomic characters and their biological features above [4,13,19]. An intraspecific genome-wide phylogeny encompassing 37 isolates, however, determined the existence of six phylogenetic lineages, which are analogous to those built earlier using only the four genes [13,17]. Despite the substantial advances as aforesaid, differentiations of mitogenomes between the isolates of *C. cassiicola* with different hosts, geographical origins, and even lifestyles have never been known.

Mitochondria are bilayer membrane organelles that arose through the phagocytic integration of an α-proteobacterial endosymbiont into the proto-eukaryotic host cell [20]. During evolution, genetic systems (mitogenomes) of the mitochondrial compartments were significantly altered, and most functions have been lost or relocated to the nuclear genomes [21]. In fungi, these "vestigial" mitogenomes generally contain a set of core protein coding genes (PCGs) involved in electron transport and coupled oxidative phosphorylation, two ribosomal RNAs (rRNAs), a varying number of transfer RNAs (tRNAs) and some unidentified open reading frames (uORFs) [22]. However, even if the genetic contents of fungal mitogenomes are relatively conserved, different fungi varied greatly in size and organization of the mitogenomes, ranging from 18.8 kb (*Hanseniaspora uvarum*) to 531.2 kb (*Morchella crassipes*) in length [23,24]. The huge variations between different fungal species mainly result from their intergenic regions and introns, which account for approximately 5% to 80% of the whole mitogenomes [23,25]. Mitochondrial introns, according to their RNA secondary structures, are classified into two types, namely group I and group II, of which the former is predominantly present in fungi and the latter dominates plant mitogenomes [26]. Group I introns can be further divided into several subgroups, such as IA1, IA3, IB, IC and ID. In addition, two families of homing endonuclease genes (HEGs), LAGLIDADG and GIY-YIG motifs, often reside in non-critical sequences of the fungal group I introns, and help their host introns transfer and proliferate across the mitogenomes [27].

Mitochondrial genomes have been widely used as effective molecular markers for inter- and intraspecific phylogenetic analyses in the fungal kingdom because of their uniparental inheritance and accelerated evolution rate [28–30]. Nonetheless, compared to the available mitogenomes of animals (>10,000), fungal mitogenomes (⟨900) are still largely underrepresented in the NCBI database (https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/), far less than the number of their nuclear genomes (>8,000). Until now, only one complete mitogenome of *C. cassiicola* isolated from kiwifruit in China has been formally published and simply described [31], while the other one (strain CCP) was just deposited in the JGI database (https://mycocosm.jgi.doe.gov/mycocosm/home). In the present study, the mitogenomes of 16*C. cassiicola* isolates collected from various host plants throughout China were sequenced, assembled, and annotated. Besides, hybrid genomic data of 38*C. cassiicola* strains distributed across the world were downloaded from public databases and used for extraction and assembly of their mtDNA fragments. Our primary objectives are: 1) to present the mitogenome characterizations of *C. cassiicola* from different host and geographical origins; 2) to reveal the intraspecific variations and conservations between these *C. cassiicola* isolates by comparative mitogenomic analysis; 3) to uncover the phylogenetic relationship within the species *C. cassiicola* based on whole mitogenomic data and intron distribution.

2. Materials and methods

2.1. Sample collection and data acquisition

Sixteen *Corynespora cassiicola* isolates were originally sampled from lesioned tissues of diverse host plants (*Fragaria* × *ananassa*, *Cucumis sativus*, *Gossypium hirsutum*, *Hevea brasiliensis*, *Lonicera maackii*, *Ligustrum quihoui*, *Phaseolus vulgaris*, *Sesamum indicum*, and *Salvia splendens*) and from different collection locations (including Henan, Shandong, Yunnan and Liaoning Provinces) in China (Supplementary Table 1). Species identification was based on the morphological features of colonies and conidia, and then confirmed using the ITS barcoding sequences. To obtain pure cultures, single-conidium isolation of each *C. cassiicola* strain was performed and then grown on the sterile potato dextrose agar (PDA) media at 25 °C (photoperiod 12 h/12 h) for 5 days. After cultivation, mycelium mats were harvested and kept in 15% aqueous glycerol, and eventually stored at − 80 °C in the fungal collection of Henan Agricultural University.

For comparative analyses, mtDNA sequence resources of another 40*C. cassiicola* strains were retrieved from two different public databases: NCBI GenBank and JGI MycoCosm portal, of which two (CCP and Kiwi) already possessed the complete mitochondrial genomes, while for all the others only hybrid (both nuclear and mitochondrial) genome assemblies were available (Supplementary Table 1). These strains were collected from diseased or asymptomatic tissues of various representative hosts (*H. brasiliensis*, *C. sativus*, *Glycine* max, *G. hirsutum*, *Solanum lycopersicum*, *Vernonia cinerea*, *Piper hispidinervum*, *Actinidia chinensis*, and *Homo sapiens*) and from most of their geographical distribution range (USA, Brazil, China, India, Sri Lanka, Malaysia, Thailand, Philippines, Cameroon, Cote d'Ivoire, and Gabon) [1,13,31,32].

2.2. Sequencing and assembly

Total genomic DNA extraction of the 16*C. cassiicola* specimens collected by us was performed using the Plant Genomic DNA Kit DP305 (Tiangen, Beijing, China) following manufacturer's instructions. Short insert library (350 bp) was constructed using the NEBNext® Ultra DNA Library Prep Kit for Illumina (NEB, USA). Whole-genome sequencing was carried out on an Illumina Hiseq X Ten platform (Novogene, Tianjin, China), generating 150 bp paired-end reads for each sample. Raw data were trimmed of adapters and low-quality reads (ratio of the bases less than Q20 > 30% or containing undetermined bases) by using the fastp v0.13.1 [33]. FastUniq v1.1 [34] and Musket v1.1 [35] were successively used for duplicate removing and error correction. Cleaned paired-end reads were *de novo* assembled using the SPAdes v3.14.1 software with kmers of 21, 33, 55, 77, 99 and 127 [36]. Mitochondria-related contigs were identified and pooled for each sample by performing BLASTN search against the mitogenome of *C. cassiicola* isolate Kiwi. Then, we filled gaps between these contigs by using the MITObim v1.9 to build closed-circular mtDNA macromolecules of the *C. cassiicola* samples [37].

As for the other 38*C. cassiicola* strains (excluding CCP and Kiwi) that had no ready-made mitogenome sequences, we extracted mtDNA fragments from their hybrid genome assemblies using the same method as above, and then tried to assemble them based on the reference mitogenome of *C. cassiicola* isolate Kiwi. However,

it is a pity that there was no online raw sequencing data available for their subsequent gap filling.

2.3. Annotation of mitogenomes

Initial annotation of the complete mitogenomes was performed with the online tools MFannot [38] and MITOS2 [39], both based on the genetic code 4. All the mitochondrial elements including protein coding genes (PCGs), unidentified open reading frames (uORFs), introns, rRNAs and tRNAs were named according to their usual nomenclature [40]. Gene boundaries were adjusted by performing BLASTN search against the database of all nonredundant GenBank sequences (nr database). The exon–intron borders of PCGs were corrected by comparing them with the corresponding intron-free genes of closely related species. Then ORFs were modified or predicted using the NCBI Open Reading Frame Finder (https://www.ncbi.nlm.nih.gov/orffinder), and further annotated by BLASTP searches (e-value < 1e-5) against the NCBI nonredundant protein sequence database. We also predicted tRNA genes in the *C. cassiicola* mitogenomes using tRNAscan-SE v2.0 [41], and detected their secondary structures using MITOS with default parameters [39]. Graphical maps of the mitochondrial genomes were drawn with the online program OGDraw v1.2 [42].

2.4. Sequence analysis

Base compositions of the *C. cassiicola* mitogenomes were calculated using the DNASTAR Lasergene v7.1 software (http://www.dnastar.com/). The CodonW v1.4.2 software (http://codonw.sourceforge.net/) was used to analyze the codon usage frequency and preference of core PCGs in the *C. cassiicola* mitogenomes. Strand asymmetries of the *C. cassiicola* mitogenomes were assessed based on the following formula: GC skew = [G - C] / [G + C], and their GC-skew plots and skew indices (SkewI) were presented using the method SkewIT with 2000 window size and 1000 frequency [43]. The pairwise genetic distances between each pair of the 13 core PCGs (*atp6*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, and *rps3*) and two rRNA genes (*rns* and *rnl*) in the *C. cassiicola* mitogenomes were calculated using MEGA v7.0.26 based on the Kimura-2-parameter (K2P) substitution model [44]. The DnaSP v6.12.03 software was used to calculate synonymous (*Ks*) and nonsynonymous substitution rates (*Ka*) for core PCGs in the *C. cassiicola* mitogenomes [45]. Whole mitogenome comparison of the *C. cassiicola* strains was performed using the BLAST Ring Image Generator (BRIG) v0.95 [46]. Toxin (cassiicolin) class for each *C. cassiicola* strain was acquired by consulting related literatures [13], or determined by BLASTN search against their nuclear genome sequences using the known *Cas* genes [2].

2.5. Dynamic analysis of group I introns

Core PCGs (such as *cox1*) of fungal mitogenomes were frequently reported to contain group I introns [40]. Of all the 56 *C. cassiicola* mitogenomes, intron-containing core PCGs were extracted and aligned with their corresponding coding sequences (CDSs) by Clustal W [47] to detect the insertion positions (IPs). Besides, group I introns of the same type (e.g., IB and ID) from the same or different IPs were aligned to determine their similarity and homology. Due to the non-uniform distribution of group I introns within some intron-containing core PCGs, the 56 *C. cassiicola* strains could be classified into different groups according to the identity and IP of the introns.

2.6. Phylogenetic inference

In order to investigate the intraspecific phylogeny of the 56 *C. cassiicola* strains, we performed phylogenetic analysis using the concatenated mitochondrial gene set (CDSs of 13 core PCGs + 2 rRNAs). Individual mitochondrial genes were first aligned using the MAFFT v7.475 with auto strategy [48], and then concatenated into a combined mitochondrial gene set using the SequenceMatrix v1.7.8 [49]. Partition homogeneity test was used to detect potential phylogenetic conflicts between different mitochondrial genes. Best-fit models of phylogeny and partitioning schemes for the gene set were determined using the PartitionFinder v2.1.1 [50]. Phylogenetic trees were constructed using both maximum likelihood (ML) and Bayesian inference (BI) methods. The ML analysis was implemented in the RAxML v8.2.12 with 1,000 bootstrap replications under the substitution model GTRGAMMA [51]. The BI analysis was performed using the MrBayes v3.2.6 [52]. Four simultaneous Markov chains were run starting from a random tree for 2,000,000 generations and trees were sampled every 100 generations. The first 25% of samples were discarded as burn-in, and the remaining trees were used to calculate values of Bayesian posterior probabilities (BPPs) in a 50% majority-rule consensus tree. Tree rendering was carried out using the FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

2.7. Data availability

The complete mitogenomes of the 16 *C. cassiicola* isolates sequenced here were deposited in the GenBank database under the accession numbers of OK054352-OK054367, and their raw sequencing data were deposited in the Sequence Read Archive (SRA) database under the accession numbers of SRR16654377-SRR16654392 (Supplementary Table 1).

3. Results

3.1. Mitogenome assemblies of 56 C. cassiicola strains

A total of 22,928,836 clean reads of the *C. cassiicola* isolate HG2 were assembled into 43.97-Mb hybrid genome assembly containing 3,697 contigs, of which only one (43,975 bp in length) was well mapped to the reference mitogenome of *C. cassiicola* isolate Kiwi. These 150-bp PE reads were then reused to cyclize the mtDNA-related contig into a complete mitochondrial genome (43,965 bp). Average coverage of the *C. cassiicola* HG2 mitogenome could reach 1330-fold (vs. 77-fold of its nuclear genome), and its GC content was 29.46%. With the similar method, intact mitogenome sequences of all 15 other *C. cassiicola* isolates collected by us were obtained (Supplementary Table 2). Their mitogenome sizes, average coverages, and GC contents ranged respectively from 40,537 bp to 43,976 bp, from 1011-fold to 3507-fold, and from 29.05% to 29.46%.

Of the 40*C. cassiicola* strains (except Kiwi and CCP) from public databases, twelve strains obtained complete circular mtDNA molecules, while 26 strains only had mitogenome-scale scaffolds with different lengths of ambiguous regions (from 25 to 1,202 non-ATCG bases). In summary, the sizes of the 40*C. cassiicola* mitogenomes ranged from 39,223 bp to 45,786 bp with an average value of 42,513 bp, and their GC contents ranged from 28.30% to 29.44% with an average value of 29.11% (Supplementary Table 2).

3.2. PCGs and rRNA genes of C. cassiicola mitogenomes

All of the 56 *C. cassiicola* mitogenomes harbored the ribosomal protein subunit 3 (*rps3*) and 12 typical fungal mitochondrial PCGs

related to the oxidative phosphorylation pathway, including apocytochrome b (*cob*), ATP synthase membrane subunit 6 (*atp6*), three cytochrome oxidase subunits (*cox1*, *cox2* and *cox3*) and seven NADH dehydrogenase subunits (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*). The 13 conserved core PCGs appeared on both strands in an unbiased fashion, and nine of them were intronless. Specifically, cox1 was seamlessly adjacent to cox2, resembling a fused gene, while the ATG initiation codon of *nad5* followed immediately after the termination codon of *nad4L*, with an overlap of one base. In addition, 2 to 4 free-standing open reading frames of unknown function (uORFs) were found in the intergenic regions of each *C. cassiicola* mitogenome, and were named after the lengths of their protein products, e. g., uORF311. Although probably with various gene sizes, the uORFs located at the similar loci of different mitogenomes were generally homologous.

Each *C. cassiicola* mtDNA had two mitochondrial ribosomal RNA (rRNA) genes, namely the small and large rRNA subunits (*rns* and *rnl*). The *rns* genes were 1,674 bp in length and the sizes of the *rnl* genes were 3,231 bp or 3,232 bp, both of which had no introns. Arrangements of the two rRNAs and above 13 core PCGs in the 56 *C. cassiicola* strains were completely consistent, following the order: *cox1*, *cox2*, *nad4L*, *nad5*, *nad1*, *cob*, *nad4*, *nad3*, *nad2*, *rps3*, *atp6*, *rns*, *nad6*, *rnl*, and *cox3*.

3.3. Distribution and alignment of group I introns

Four mitochondrial core PCGs (*cox1*, *cox3*, *nad1*, and *nad5*) were predicted to contain a variety of group I introns across the 56 *C. cassiicola* strains. Within the CDSs (1,644 bp in length) of *cox1* genes, three group IB introns (1,287 bp, 1,316 bp, and 1,174 bp) were individually or simultaneously inserted into different positions (P213, P387, and P616), and were named Cox1.BP213, Cox1.BP387 and Cox1.BP616, respectively. Cox1.BP213 carried one ORF encoding GIY-YIG homing endonuclease (HE), while both Cox1.BP387 and Cox1.BP616 harbored LAGLIDADG HE genes. In total, there were twelve Cox1.BP387-containing strains, 32 Cox1.BP616-containing strains, three Cox1.BP213&Cox1.BP387-containing strains, two Cox1.BP387&Cox1.BP616-containing strains, and seven intronless-*cox1*-containing strains. Analogously for the 1,980-bp CDSs of *nad5* genes, one group IC2 intron (1,351–1,357 bp) and two group IB introns (1,013 bp and 1,876 bp) were individually or simultaneously inserted into different positions (P325, P718, and P925). They all contained ORFs encoding LAGLIDADG HEs and were named Nad5.C2P325, Nad5.BP718 and Nad5.BP925, respectively. In total, there were 48 Nad5.BP718-containing strains, five Nad5.C2P325&Nad5.BP718-containing strains, and three Nad5.BP718&Nad5.BP925-containing strains. In addition, the 810-bp CDSs of *cox3* genes in all *C. cassiicola* strains were interrupted by one group ID intron (597–632 bp) at position 429, named Cox3.DP429. The 1,116-bp CDSs of *nad1* genes in 35*C. cassiicola* strains were interrupted by one group IC1 intron (1,499–1,507 bp) at position 973, named Nad1.C1P973. Neither of them was identified to encode any HEs.

Multiple sequence alignment analysis showed that there were no similarities between the above eight group I introns, while no significant differences were found among these introns from the same insertion site. Both Cox1.BP213s and Nad5.BP925s had absolutely identical sequences; only 32 SNPs were present among the Cox1.BP387s; eight SNPs among the Cox1.BP616s; 17 SNPs (including nine indels) among the Nad5.C2P325s; 17 SNPs among the Nad5.BP718s; 81 SNPs (including 45 indels) among the Cox3.DP429s; and 50 SNPs (including 15 indels) among the Nad1.C1P973s (Supplementary Fig. 1). Furthermore, BLASTN search results indicated that all these group I introns shared high coverages and sequence identities with corresponding introns in the mitogenomes of fungi outside the genus *Corynespora*, except Nad5.BP925 and Cox3.DP429 that had only vestigial homologs (Supplementary Table 3).

3.4. Grouping of the C. cassiicola strains based on both intron distribution and phylogeny

According to the differential distribution of group I introns of the *cox1*, *nad1* and *nad5* genes, the 56 *C. cassiicola* strains could be divided into eight groups named Group1 to Group8 (Fig. 1A). Group1 covered 23 strains harboring Cox1.BP616, Nad5.BP718 and Nad1.C1P973; Group2 covered nine strains harboring Cox1.BP387, Nad5.BP718 and Nad1.C1P973; Group3 covered three strains harboring Cox1.BP213, Cox1.BP387, Nad5.BP718, Nad5.BP925 and Nad1.C1P973; Group4 covered five strains harboring Cox1.BP616, Nad5.C2P325 and Nad5.BP718; Group5 covered four strains harboring Cox1.BP616 and Nad5.BP718; Group6 covered two strains harboring Cox1.BP387, Cox1.BP616 and Nad5.BP718; Group7 covered three strains harboring Cox1.BP387 and Nad5.BP718; Group8 covered seven strains harboring only Nad5.BP718.

In addition, we obtained tree topologies using both the ML and BI methods based on the concatenated alignment of 15 gene loci (13 core PCGs + 2 rRNAs). This analysis involved 56 nucleotide sequences containing 19,939 positions (including gaps), of which 242 characters were variable and 235 were parsimony informative. As no obvious differences were found between the ML and BI trees, only the former was displayed here with both bootstrap values and BPPs indicated at the nodes (Fig. 1B). From the intraspecific mitogenome-wide phylogeny of 56 *C. cassiicola* strains, eight well supported clades could be easily distinguished, exactly corresponding to the above eight groups established based on intron distribution. Except for Group8, all the groups were monophyletic, and within part of them, some highly-supported subclades were detected.

As the biggest population, 23 members of Group1 had the most complicated sources, sampled from a variety of host plants (*H. brasiliensis*, *C. sativus*, *F.* × *ananassa*, etc.) in China, Brazil, Southeast Asia (Thailand and Malaysia), West Africa (Gabon and Côte d'Ivoire) and South Asia (India and Sri Lanka), respectively. Their toxin isoforms included Cas0, Cas2 and Cas2 + 6 (Fig. 2). Group2 consisted of nine *C. cassiicola* isolates (involved in Cas0, Cas2, Cas1 + 2 and Cas2 + 7) from *G. hirsutum*, *S. indicum* and *P. vulgaris* in China, *C. sativus* and *V. cinerea* in Brazil, and *G. hirsutum* in the USA, respectively. Group3 was composed of three Cas0 isolates collected on *H. brasiliensis* and *P. hispidinervum* in Brazil and *S. lycopersicum* in the USA. Group4 contained five isolates from *H. brasiliensis* forming three deep-branching subclades: two Cas0 isolates of Thailand, one Cas0 isolate of China, and two Cas4 isolates of Brazil. Group5 comprised three Cas1 isolates from *H. brasiliensis* in West Africa (Cameroon) and Southeast Asia (the Philippines), and one kiwi isolate from China (unknown toxin class without nuclear genomic data). Group6, as the simplest clade, was represented by two Cas0 isolates collected from *H. brasiliensis* in Cameroon. Group7 contained two Cas2 isolates from *S. splendens* in China and one Cas2 isolate from *G.* max in Brazil. Finally, Group8 was polyphyletic, including one Cas0 isolate from contact lens of a patient in Malaysia and a high-supported clade involving six isolates (Cas0 and Cas5) collected on *H. brasiliensis* from Malaysia, Thailand and Sri Lanka.

3.5. Mitogenome-wide comparative analysis

Besides the high degree of sequence similarity, *C. cassiicola* mitogenomes within the same group also had similar assembly sizes, GC contents and genetic structures. Therefore, one strain with an ungapped circular mtDNA molecule from each group (except Group6) was picked out to represent its whole group (Fig. 3)

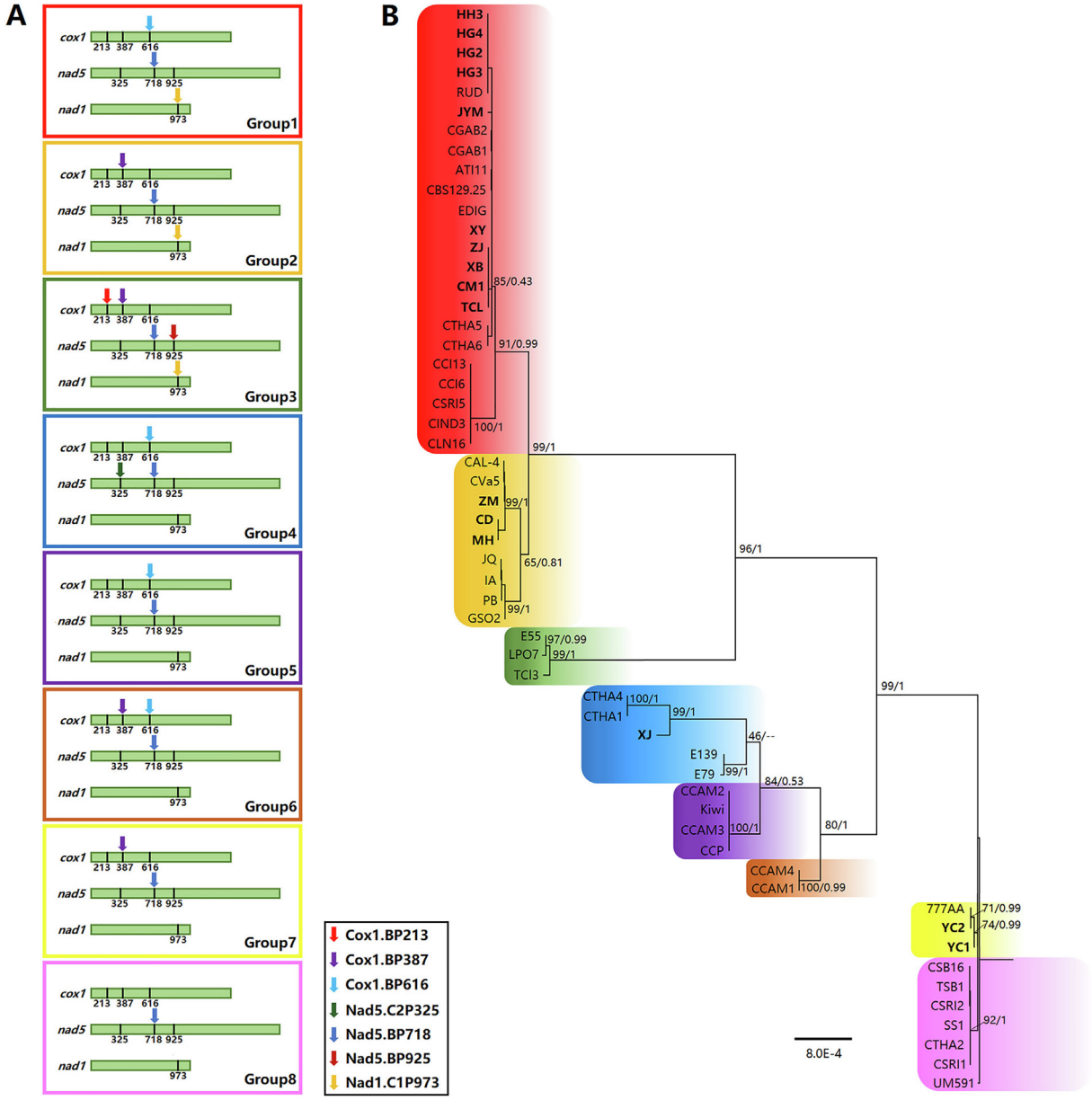


Fig. 1. Grouping analysis of 56 *Corynespora cassiicola* isolates based on both intron distribution and molecular phylogeny. (A) Eight *C. cassiicola* groups established through dynamic analysis of seven group I introns inserted in three core protein-coding genes (PCGs) *cox1*, *nad5* and *nad1*. (B) Phylogenetic tree constructed based on Maximum Likelihood (ML) and Bayesian inference (BI) methods using the 13 core PCGs and two rRNAs. Bootstrap (BS) values and Bayesian posterior probabilities (BPPs) are repectively placed before and after the slash. Coloured clades exactly corresponds to the different groups shown in Fig. 1A.

and to be used for subsequent comparative analysis. Firstly, BLAST comparisons were carried out by using the longest mitogenome sequence (from strain TCl3 of Group3) as reference and the other seven representative mitogenomes as queries (Fig. 4). The BRIG result revealed that the main genetic differences among the eight groups appeared to be simply the insertions and deletions (indels) of group I introns within the three genes *cox1*, *nad1* and *nad5*. Besides, a few minor structural variations (indels) were located in the intergenic regions between *nad5* and *nad1* and between *nad4* and *nad3*, respectively. Secondly, GC skew plots were generated by calculating GC skews in adjacent or overlapping windows across the full lengths of all the eight representative mitogenomes. Analysis of these plots indicated that the eight groups displayed similar but not identical fluctuation trends of the GC skew values along their mitogenomes (Supplementary Fig. 2). To sum up, the regions (from *rns* to *nad5*) rich in sense strands generally presented positive GC skews, while the regions (from *nad1* to *atp6*) corresponding to antisense strands often presented negative GC skews. All the eight representative strains had strong GC skew patterns, with the SkewI values equal to 1.

3.6. tRNA genes and codon usage analysis

Each *C. cassiicola* mitogenome carried the same set of 27 tRNA genes with lengths ranging from 71 bp to 85 bp. These tRNAs encoded 20 standard amino acids, including two copies of *trn*I (for isoleucine, Ile), two copies of *trn*L (for leucine, Leu), three copies of *trn*M (for methionine, Met), two copies of *trn*R (for arginine, Arg), two copies of *trn*S (for serine, Ser), two copies of *trn*V (for valine, Val), and the other 14 single-copied tRNA genes. Most (21/27) of the tRNA genes clustered in the region of *nad6*-(10 tRNAs)-*rnl*-(11 tRNAs)-*cox3*. Among the eight groups of *C. cassiicola*

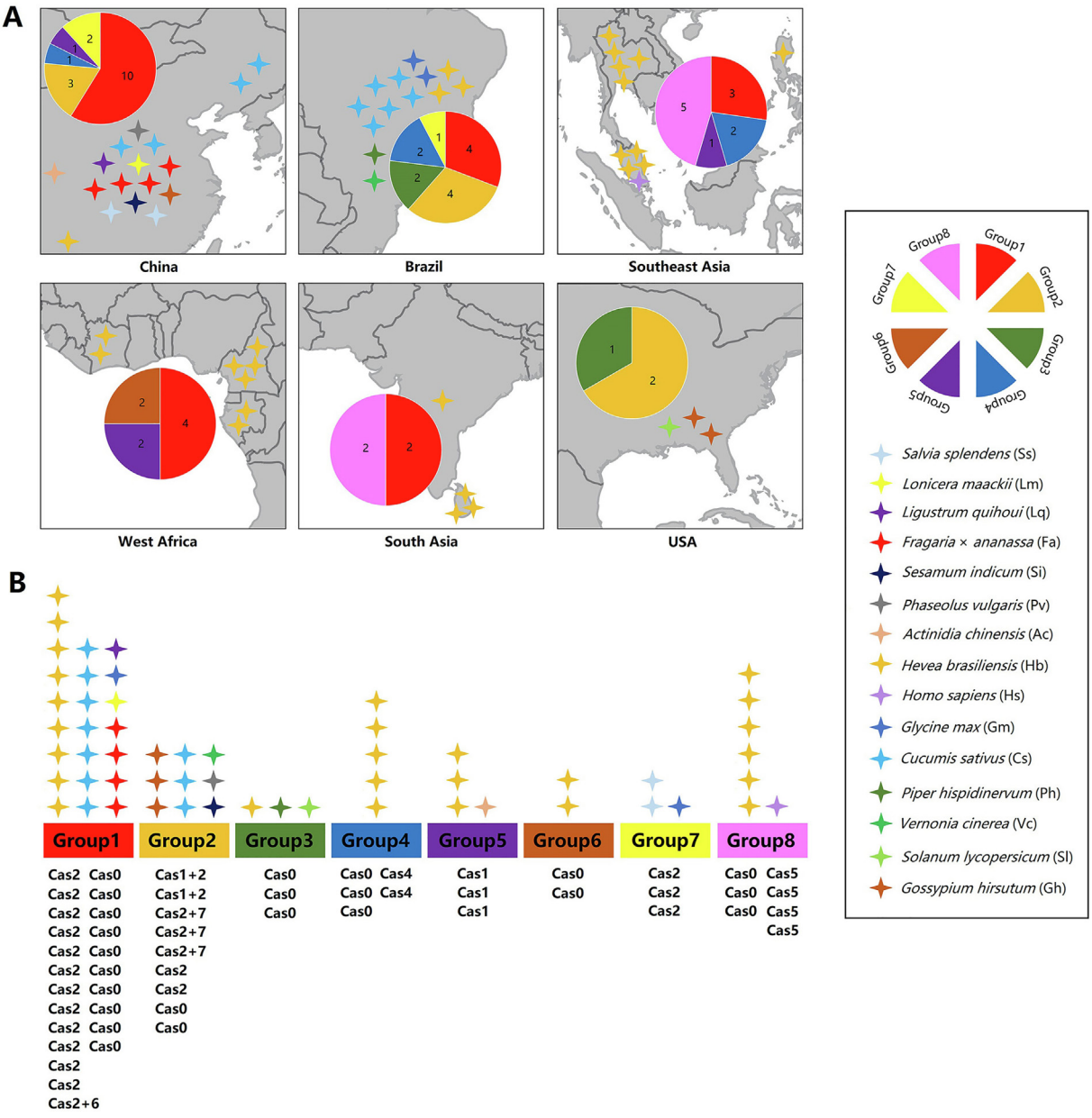


Fig. 2. Relationship of the eight *Corynespora cassiicola* groups to geographical origin, host range and toxin (cassiicolin) class. (A) Distribution of the eight *C. cassiicola* groups and their host sources in six different geographical area worldwide. (B) Statistics of the number of *C. cassiicola* isolates with various host sources (above) and toxin classes (below) in each group.

strains, 21 tRNA sequences were identical, whereas six tRNA genes were found to have eight variable sites (Fig. 5). In Group5 and Group6, P57 of the *trn*H (anticodon: gug) gene was adenine nucleotide (A) instead of guanine nucleotide (G) in the other groups; from Group4 to Group6, P19 of *trn*I (uau) was uracil nucleotide (U) instead of G; in Group7 and Group8, P61 of *trn*I (uau) was G instead of A; in Group5, P47 of *trn*P (ugg) was U instead of A; in Group1 and Group2, P57 of *trn*R (ucu) was A instead of U; distinguishingly, mutations of P16 of *trn*I (uau), P9 of *trn*M (cau)-1 and P28 of *trn*S (uga) occurred only between some strains of Group4.

Of the 13 core PCGs in 56 *C. cassiicola* mitogenomes, five used UAG as the stop codon and eight used UAA as the stop codon. The AUG was the only start codon used in the mitogenomes of *C. cassiicola*. For most of the 64 triplet codons, different *C. cassiicola* groups had slightly distinct frequencies of utilization, while for some codons, even within the same group, their usage frequencies could be also different (Supplementary Table 4). Codon usage analysis indicated that the most frequently used codons were UUA (for Leu), UUU (for phenylalanine, Phe), AUA (for Ile), AUU (for Ile), AAU (for asparagine, Asn), and UAU (for tyrosine, Tyr). However, over half of them had no corresponding anticodons from the above 27 mitochondrial tRNAs. The frequent use of A and U in the codons contributed to the high AT content (70.81% in average) of the 56 *C. cassiicola* mitogenomes.

3.7. Variations and evolutionary rates of the core PCGs and rRNAs

Across the 13 core PCGs and two rRNAs, sequence lengths of four genes (*cox1*, *cox3*, *nad1*, and *nad5*) varied dramatically among the eight *C. cassiicola* groups (Fig. 6A). The *cox1* gene had the largest variation degree in length (from 1,644 bp to 4,247 bp), followed by *nad5* (2,993 bp to 4,869 bp) and *nad1* (1,116 bp to 2,621 bp),

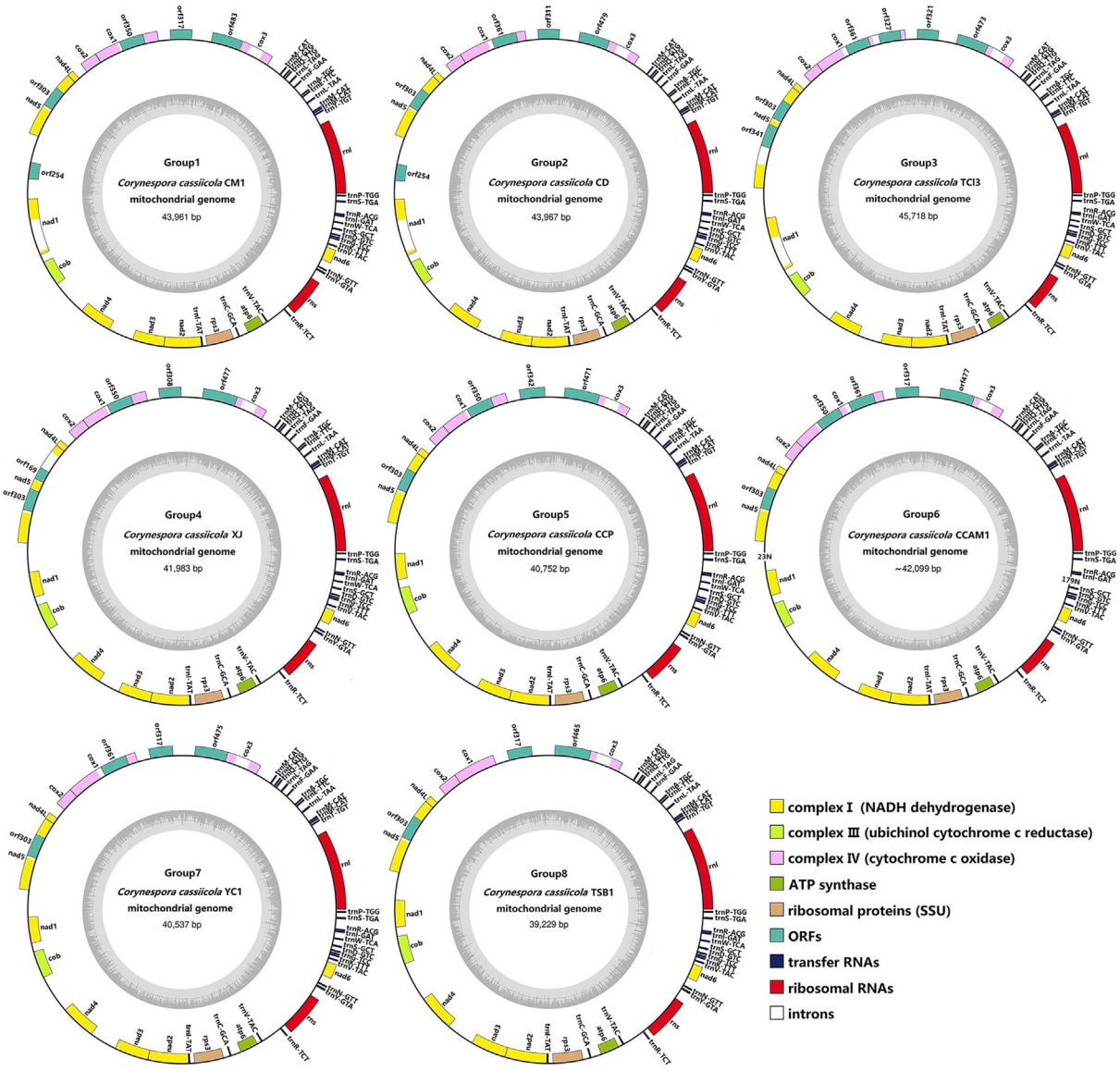


Fig. 3. Circular maps of eight representative mitogenomes from the different *Corynespora cassiicola* groups. Colored blocks along the outer rings represent the mitochondrial core protein-coding genes, rRNAs, tRNAs, and ncORFs. Inner circles show the GC content. All these circular mtDNA molecules are intact, except CCAM1 that has two gaps containing 23 and 179 non-ATCG bases, respectively.

whereas the *cox3* gene varied the least (1,408 bp to 1,440 bp). Analogously, GC contents of the four genes *cox1*, *cox3*, *nad1* and *nad5* varied significantly among the eight groups (Fig. 6B). The *cox1* had the largest variation degree (29.60%–32.60%), followed successively by *nad1* (28.30%-30.40%), *nad5* (28.00%-30.00%), and *cox3* (33.40%-34.10%). Of the 15 genes detected, *rns* contained the highest GC content (36.10% in average), and *nad6* had the lowest GC content (22.60% in average).

Across the 13 core PCGs and two rRNAs examined, *nad3* exhibited the greatest K2P genetic distance (mean value 0.0159) among the 56 *C. cassiicola* strains, followed by *cox1* (mean value 0.0105), which indicated that they had the fastest mutation rates (Fig. 7). The *nad6* and *rps3* exhibited the least genetic differentiations, with an overall mean K2P distance of 0.0003 and 0.0006, respectively, indicating that the two genes were highly conserved. In addition, the *nad3* gene had the highest mean nonsynonymous substitution rate (*Ka*, 0.0118) among the 56*C. cassiicola* strains, while the genes *cox2*, *nad4L* and *nad6* had the lowest mean *Ka* values (all equal to 0) (Fig. 7). The *cox1* gene had the highest mean synonymous substitution rate (*Ks*, 0.0400) while the *rps3* gene had the lowest mean *Ks* value (0.0004). The *Ka*/*Ks* values for most of the core PCGs were less than 1 (Fig. 7), suggesting that these genes had been subjected to purifying selection in the microevolution process of *C. cassiicola*. Nevertheless, the *Ka*/*Ks* value of *nad3* was observed to be more than 1 between some *C. cassiicola* strains from Group4 to Group8, indicating that this gene may be under pressure of positive selection in these populations.

4. Discussion

Mitochondrial genomes are assumed to have undergone the dramatic contractions of gene contents since the α-proteobacteria resided in proto-eukaryotic cells as endosymbionts [21]. In typical fungal mitogenomes, only dozens of conserved functional genes including the core PCGs (involved in respiration and protein synthesis) and noncoding RNAs (rRNAs and tRNAs) are retained. Besides, varying quantities of introns, uORFs, repeti-

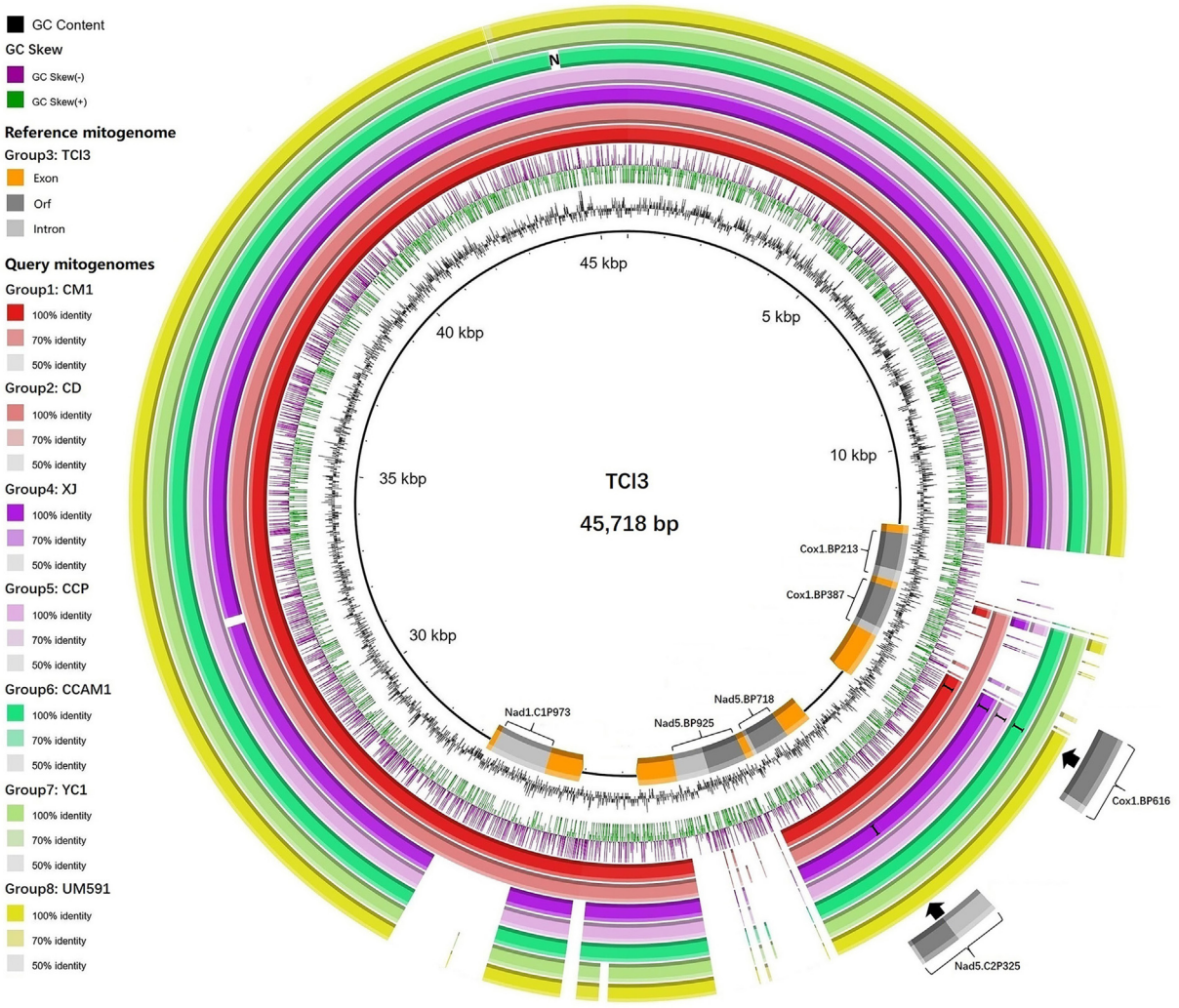
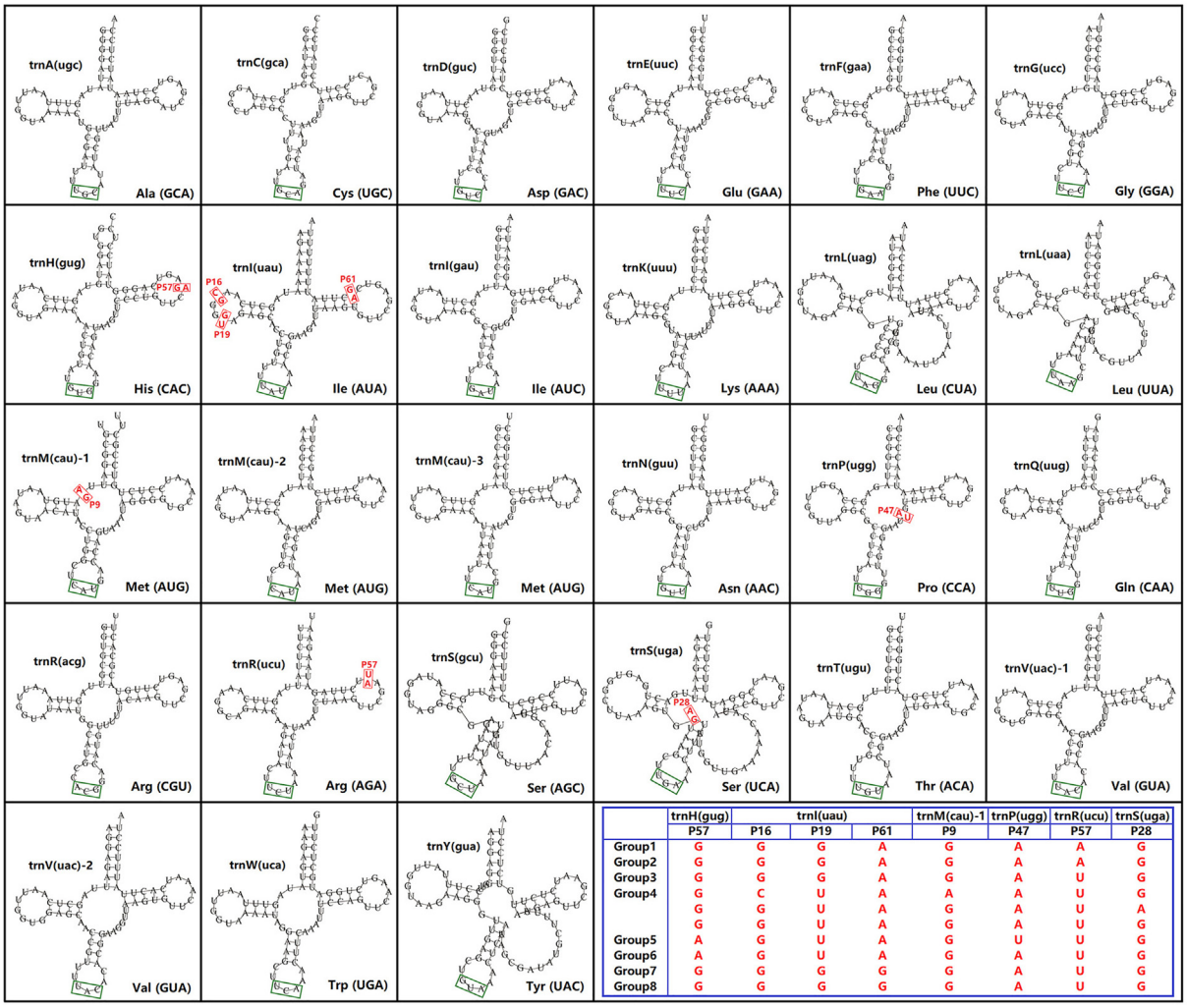


Fig. 4. BLAST ring comparisons of seven other representative mitogenomes against the longest (reference) mitogenome of *Corynespora cassiicola* TCl3. Major structure variations caused by indels of introns within the query mitogenomes are indicated by the outermost and innermost grey rectangles.

tive elements or plasmid-related DNA fragments are generally found, which largely leads to the expansions of fungal mitogenomes [22]. In the present study, a total of 56 mitogenomes of *C. cassiicola* were investigated and analyzed comprehensively, which ranged from 39,223 bp to 45,786 bp in size, with an average value of 42,770 bp. Compared with the 14 complete mitogenomes from the order Pleosporales published in the NCBI Organelle Genome database, *C. cassiicola* mtDNA molecules are just slightly larger than that of *Shiraia bambusicola* (39,030 bp), but much smaller than the biggest one from *Pyrenophora tritici-repentis* (237,114 bp). This huge variation is mainly associated with the number and length of introns as well as uORFs, for the mitogenomes of Pleosporales always have similar gene compositions and relatively few repeat structures [53,54]. For instance, in the 98,533-bp mitogenome of *Coniothyrium glycines*, 32 introns comprising 54.1% of the total mtDNA were identified [55], while in the mitogenome (137,775 bp) of *Bipolaris sorokiniana*, a total of 28 introns and 52 uORFs were detected [56]. All these figures far exceed the numbers of introns (2–6) and uORFs (2–4) in *C. cassiicola* isolates, occupying 9.8–21.6% of their mitogenomes.

As mentioned above, the gene content of the Pleosporales mitogenomes is quite conserved, consisting mainly of *nad1-nad6*, *nad4L*, *cob*, *cox1-cox3* and *atp6* [53]. Another two core PCGs, *atp8* and *atp9*, are absent from the sequenced mitogenomes of Pleosporales (including *C. cassiicola*) [57], which however could be commonly observed in other orders of Dothideomycetes, such as Capnodiales and Dothideales [58,59]. Unlike in *Stemphylium lycopersici* [53], we did not find their homologues in the nuclear genomes of *C. cassiicola*, suggesting that these two genes might be not essential for ATP synthesis in mitochondria and were lost at some point after birth of the order Pleosporales. Additionally, a notable feature of the *C. cassiicola* mitogenomes is the gene fusion between *cox1* and *cox2*, that is, the two genes are encoded by a single continuous ORF without a canonical stop codon between them. This peculiar gene architecture was also found in some other species, such as *C. glycines* and *S. lycopersici*, but rare in the fungi outside the order Pleosporales [53,55]. In comparison to the high conservation of gene content, a remarkable variation in gene order and orientation was observed between different members of Pleosporales [55]. Indeed, the mitochondrial gene arrangement of *C. cassiicola* differed considerably from other species of this order, and only several conserved gene blocks were found, like *nad2-nad3* and *nad4L-nad5* (Supplementary Fig. 3). Also, *cox1-cox2* is an uninterrupted gene pair only within the Pleosporales, though here they are fused. Additionally, another conserved pattern of ascomycetes, large tRNA gene clusters around the *rnl*, appeared in *C. cassiicola*



	trnH(gug)	trnI(uau)			trnM(cau)-1	trnP(ugg)	trnR(ucu)	trnS(uga)
	P57	P16	P19	P61	P9	P47	P57	P28
Group1	G	G	G	A	G	A	A	G
Group2	G	G	G	A	G	A	A	G
Group3	G	G	G	A	G	A	U	G
Group4	G	C	U	A	A	A	U	G
	G	G	U	A	G	A	U	A
	G	G	U	A	G	A	U	G
Group5	A	G	U	A	G	U	U	G
Group6	A	G	U	A	G	A	U	G
Group7	G	G	G	G	G	A	U	G
Group8	G	G	G	G	G	A	U	G

Fig. 5. Putative secondary structures of tRNA genes in the 56 *Corynespora cassiicola* mitogenomes. The anticodons are indicated in green rectangles. Residues conserved across all mitogenomes are shown in black, while variable sites are shown in red with detailed information listed in the bottom-right table. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

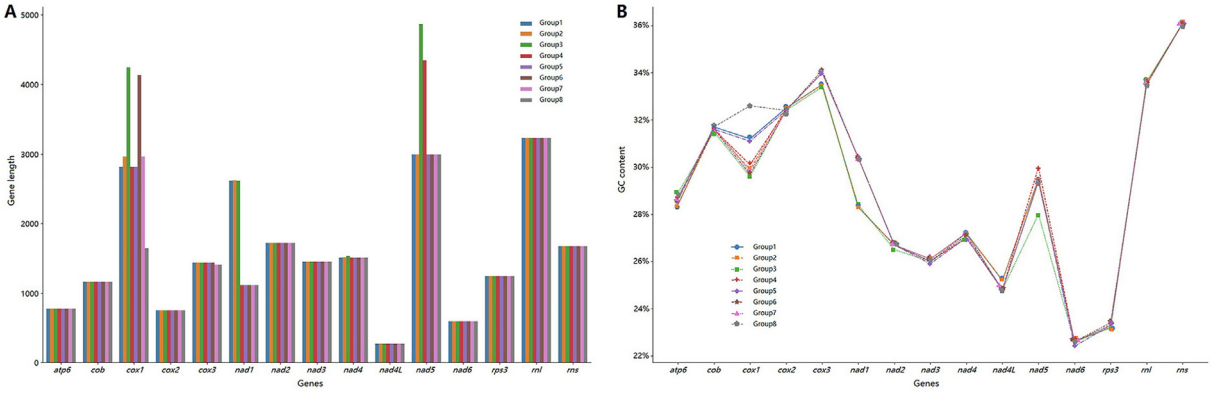


Fig. 6. Variations in the length (A) and base composition (B) of 13 core protein-coding genes (PCGs) and two rRNAs in the eight groups of *Corynespora cassiicola* mitogenomes.

[57]. This comparison of gene arrangement in mitogenomes may be used as a reference to trace the evolutionary route of fungal taxa.

Intraspecific variation of the mitogenome size of *C. cassiicola* was also conspicuous with a rangeability up to 16.7% ((max–min)/min), which even overtakes the distinction between many related species, such as *Bipolaris* spp. [56]. Comparative mitogenomic analysis showed that most (>90%) of the size variation within *C. cassiicola* was contributed by the differentially distributed group I introns located in three conserved genes, *cox1* (3 IPs), *nad1* (1 IP) and *nad5* (3 IPs). This type of introns is considered an important mobile element in the mitogenomes of fungi, and its dynamic

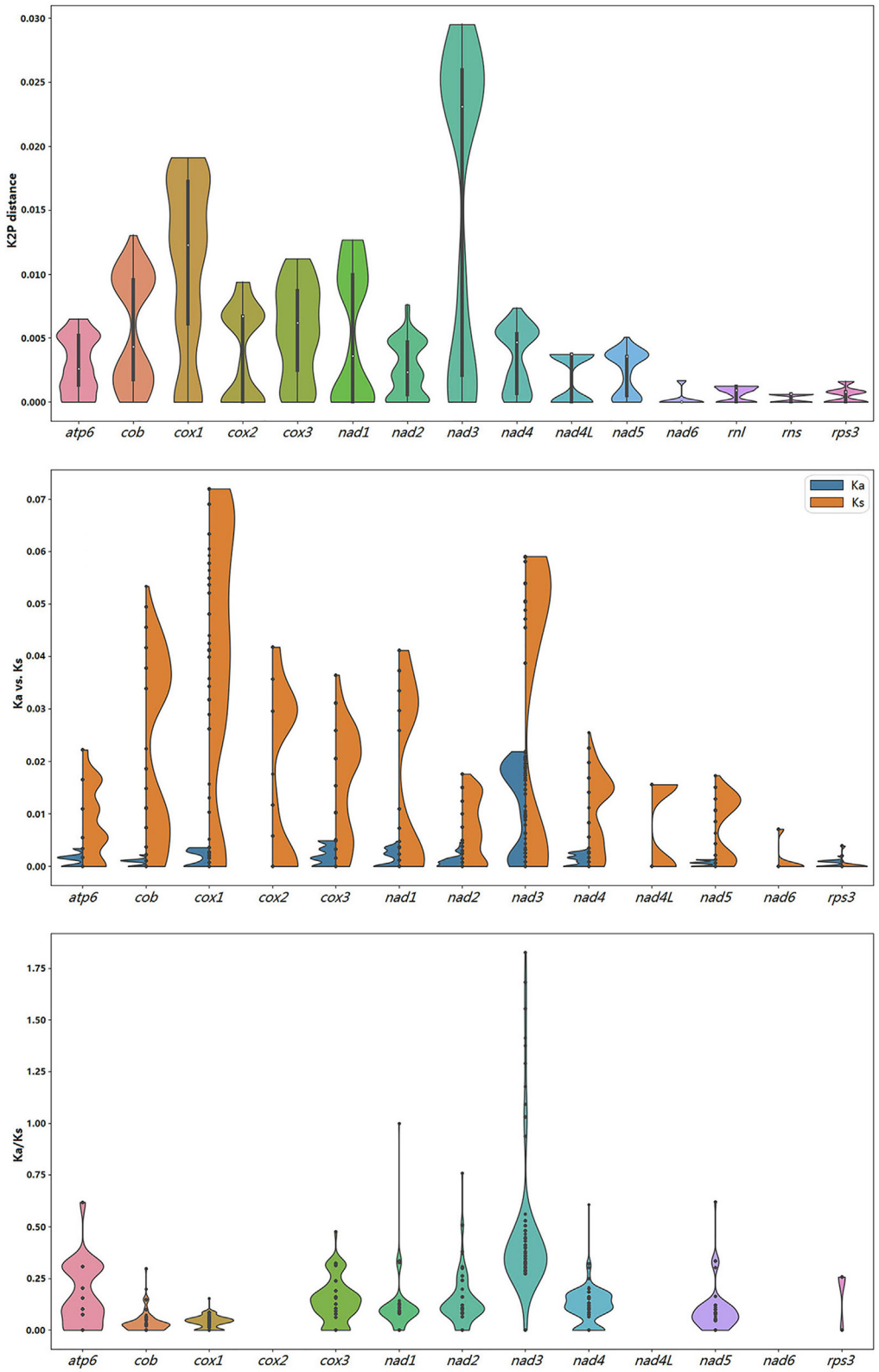


Fig. 7. Evolutionary analyses of the 13 core protein-coding genes (PCGs) and two rRNAs in the 56 *Corynespora cassiicola* mitogenomes. K2P, the Kimura-2-parameter distance; *Ka*, nonsynonymous substitution site; *Ks*, synonymous substitution site.

changes can significantly affect the size and organization of fungal mitogenomes [60]. Mobility of group I introns derives from the activity of HEs, mediating horizontal DNA transfer of the host introns even between phylogenetically distant species from different kingdom [61]. Our study indicated that three introns, Cox1. BP213, Nad5.BP718 and Nad1.C1P973, had highly homologous counterparts in the two other Pleosporales species, suggesting vertically inherited from common ancestral origin. The remaining five

introns showed high nucleotide identities to those from the fungi within other classes, such as Eurotiomycetes and Sordariomycetes. This similarity to group I introns from more distantly related species implies possible acquisition through lateral transfer rather than from a common ancestor. Regardless of their modes of transmission, the eight group I introns may have experienced discrepant gain/loss events along with the process of genetic differentiation within *C. cassiicola*, resulting in the intraspecific grouping based on intron distribution.

C. cassiicola is universally recognized as a fungal species with rich intraspecific genetic diversity, which to some extent correlates with host specialization, growth rate, or geographic distribution [17]. Our analysis of genetic diversity within *C. cassiicola*, based on combined mitochondrial gene set and group I intron dynamics, revealed eight distinct phylogenetic groups that were partly consistent with those previously stated by different authors using various neutral genetic markers [13,15,18]. The biggest phylogenetic clade (Clade A) described by Déon et al. (2014) was here broken into three groups: Group1, Group2 and Group7. Moreover, Clade B and C corresponds to Group8 and Group5, respectively; part of Group3 is included in Clade G; isolates of Group4 are assigned to Clade D and F; Group6 is adjacent to Clade D but has no corresponding clade in Déon's study. This correspondence was basically confirmed by a subsequent genome-wide intraspecific phylogeny as well [13]. In addition, a similar multi-locus analysis, in which lots of Chinese isolates were added, proposed six phylogenetic lineages (PLs), of which PL1 includes Group1, Group2 and Group7, PL2 corresponds to Group8, PL4 includes Group5, Group6 and part of Group4, PL5 includes another part of Group4, and PL6 corresponds to Group3 [2]. Through the above comparisons, some major discordances were clearly observed and listed as follows: 1) Group1, Group2 and Group7 from our analysis were always separated ambiguously or even intermixed in other studies; 2) part of Group4 frequently clustered with Group6; 3) the human-hosted isolate UM591 was here sorted into Group8 according to its intron distribution. All these discrepancies may be rooted in the different genetic characteristics of mitogenomes (such as rapid evolution rate) from nuclear genomes and other related molecular markers. Nevertheless, we found, in agreement with previous studies [17,18,62], that the eight phylogenetic groups had weak correlations with host range and toxin class, but no relationship with geographic origin. Group4-6 and Group8 consisted almost exclusively of *C. cassiicola* isolates from rubber tree, while isolates from a wide variety of plants were distributed in Group1-3 and Group7. Additionally, Cas0 isolates nearly spread all over the phylogenetic tree, whereas some toxin classes were monopolized by only one group, such as Cas1 in Group5 and Cas5 in Group8. However, we did not find the situation that a single phylogenetic group occupied a certain region; instead, the population diversity increased with the increase of sample size within a geographical area. These results suggest that host preferences and virulence genes may have played a certain role in the biology and evolution of *C. cassiicola*, while the recent global movement of this fungus suppresses its genetic differentiation caused by geographical isolation.

Despite the high variability of mitogenomic organization in Dothideomycetes (even between closely related species) [55,57], diverse *C. cassiicola* isolates shared the same set of core PCGs, rRNAs and tRNAs in the identical arrangement, implying that this consistency could be expected at least within species. Due to the varied life styles and host preferences of *C. cassiicola* [13], these highly conserved genes however underwent a series of point mutations along their coding regions. The *nad3* was found to have the greatest gene differentiation, and it was also the only one that may be subjected to the pressure of positive selection ($Ka/Ks > 1$) between some *C. cassiicola* isolates. *Nad3* is well known to encode a pivotal subunit of complex I, the first step in the electron transport system of mitochondrial oxidative phosphorylation. In humans, SNPs of this gene could even increase the susceptibility to gastric cancer development [63]. In view of its biological significance, the polymorphism of *nad3* here was speculated to be mostly associated with the various habitat adaptations of *C. cassiicola*. To make intraspecific comparison easier, we selected one representative mitogenome from each phylogenetic group (members inside are always more conserved), and found that there were obvious differences in both length and GC content of some genes between different groups. The distinctions were resulted from the uneven distribution of group I introns as previously mentioned, which was also observed in other fungal species (e.g., *Fusarium asiaticum*) and used as markers for population genetic analysis [64]. In addition, codon usage frequency and tRNA sequences varied among the eight groups of mitogenomes, which may affect protein synthesis [65]. Deserved to be mentioned, these variations not only occurred between groups but suggested a degree of intragroup differentiation, especially for Group4, as shown in the earlier phylogeny.

5. Conclusion

In this study, a comparative analysis of 56 mitogenomes of *C. cassiicola* from diverse hosts and geographic origins was performed. Alongside two intact mitogenome sequences (CCP and Kiwi) available online, we also contributed 16 newly sequenced mitogenomes of *C. cassiicola* collected from China, and completed the first assembling of the mtDNA-related contigs from 38 previously reported isolates. Across these different *C. cassiicola* isolates, we identified the same set of 13 core PCGs, two rRNAs and 27 tRNAs arranged in the same order and orientation. However, seven group I (IB, IC1, and IC2) introns were differentially inserted in three conserved genes *cox1*, *nad1* and *nad5*, leading to varied mitogenome sizes from 39,223 bp to 45,786 bp. Combined with the intron dynamic distribution, a mitogenome-wide intraspecific phylogeny divided the 56 *C. cassiicola* isolates into eight phylogenetic groups. Isolates from different groups exhibited obvious differences in both length and GC content of some genes and a degree of variance in codon usage and tRNA sequence. In addition, the *nad3* gene was found to be subjected to the pressure of positive selection between some *C. cassiicola* isolates. These data provide us new insights for mitogenomes as reliable molecular markers being applied to the analysis of intraspecific diversity of fungi, including *C. cassiicola* and other species rich in genetic variations.

Funding

This study was supported by the Science and Technology Key Project of Henan Province (192102110164), China and Natural Science Foundation of Henan Province (202300410200), China.

CRediT authorship contribution statement

Qingzhou Ma: Data curation, Formal analysis, Visualization, Supervision. **Haiyan Wu:** Data curation, Formal analysis, Project administration, Visualization. **Yuehua Geng:** Data curation, Formal analysis, Project administration. **Qiang Li:** Writing – review & editing. **Rui Zang:** Data curation, Formal analysis, Project administration. **Yashuang Guo:** Data curation, Formal analysis, Project administration. **Chao Xu:** Funding acquisition, Writing – original draft. **Meng Zhang:** Funding acquisition, Writing – review & editing.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.11.002.

References

[1] Sumabat LG, Kemerait RC, Kim DK, Mehta YR, Brewer MT, Carter DA. Clonality and geographic structure of host-specialized populations of *Corynespora cassiicola* causing emerging target spot epidemics in the southeastern United States. PLoS ONE 2018;13(10):e0205849.

[2] Wu J, Xie X, Shi Y, Chai A, Wang Qi, Li B. Variation of cassiicolin genes among Chinese isolates of *Corynespora cassiicola*. J Microbiol 2018;56(9):634–47.

[3] Gajbhiye M, Kapadnis B. First report of *Corynespora cassiicola* causing fruit rot of pomegranate in India, its morphological and molecular characterization. Natl Acad Sci Lett 2019;42(3):253–7.

[4] Gao S, Zeng R, Xu L, Song Z, Gao P, Dai F. Genome sequence and spore germination-associated transcriptome analysis of *Corynespora cassiicola* from cucumber. BMC Microbiol 2020;20(1). https://doi.org/10.1186/s12866-020-01873-w.

[5] N. Rondon M, Lawrence K. The fungal pathogen *Corynespora cassiicola*: a review and insights for target spot management on cotton and Soya bean. J Phytopathol 2021;169(6):329–38.

[6] Lee S, Melnik V, Taylor J, Crous P. Diversity of saprobic hyphomycetes on Proteaceae and Restionaceae from South Africa. Fungal Divers 2004;17:91–114.

[7] Cai L, Ji K-F, Hyde KD. Variation between freshwater and terrestrial fungal communities on decaying bamboo culms. Antonie Van Leeuwenhoek 2006;89(2):293–301.

[8] Promputtha I, Lumyong S, Dhanasekaran V, McKenzie EHC, Hyde KD, Jeewon R. A phylogenetic evaluation of whether endophytes become saprotrophs at host senescence. Microb Ecol 2007;53(4):579–90.

[9] Suryanarayanan TS, Murali TS, Thirunavukkarasu N, Govinda Rajulu MB, Venkatesan G, Sukumar R. Endophytic fungal communities in woody perennials of three tropical forest types of the Western Ghats, southern India. Biodivers Conserv 2011;20(5):913–28.

[10] Déon M, Scomparin A, Tixier A, Mattos CRR, Leroy T, Seguin M, et al. First characterization of endophytic *Corynespora cassiicola* isolates with variant cassiicolin genes recovered from rubber trees in Brazil. Fungal Divers 2012;54(1):87–99.

[11] Xie Z, Wu W, Meng D, Zhang Q, Ma Y, Liu W, et al. A case of Phaeohyphomycosis caused by *Corynespora cassiicola* infection. BMC Infect Dis 2018;18(1). https://doi.org/10.1186/s12879-018-3342-z.

[12] Luo H, Tao M, Zhang J, Cao J, Hong H, et al. A case of phaeohyphomycosis caused by *Corynespora cassiicola*, a plant pathogen. J Invest Dermatol 2020;140(7):S43.

[13] Lopez D, Ribeiro S, Label P, Fumanal B, Venisse J-S, Kohler A, et al. Genome-wide analysis of *Corynespora cassiicola* leaf fall disease putative effectors. Front Microbiol 2018;9:276.

[14] Qi Y-X, Zhang X, Pu J-J, Liu X-M, Lu Y, Zhang He, et al. Morphological and molecular analysis of genetic variability within isolates of *Corynespora cassiicola* from different hosts. Eur J Plant Pathol 2011;130(1):83–95.

[15] Wu J, Xie X, Shi Y, Chai A, Wang Qi, Li B. Analysis of pathogenic and genetic variability of *Corynespora cassiicola* based on iPBS retrotransposons. Can J Plant Pathol 2019;41(1):76–86.

[16] Hieu ND, Nghia NA, Uyen NTK, Chau NNB, Quoc NB. Genetic diversity analysis of *Corynespora cassiicola* isolates on the rubber tree (*Hevea brasiliensis*) in Vietnam using ribosomal DNA internal transcribed spacer (rDNA-ITS) sequences and sequence related amplified polymorphism (SRAP). J Rubber Res 2020;23(3):173–85.

[17] Dixon LJ, Schlub RL, Pernezny K, Datnoff LE. Host specialization and phylogenetic diversity of *Corynespora cassiicola*. Phytopathology 2009;99(9):1015–27.

[18] Déon M, Fumanal B, Gimenez S, Bieysse D, Oliveira RR, Shuib SS, et al. Diversity of the cassiicolin gene in *Corynespora cassiicola* and relation with the pathogenicity in *Hevea brasiliensis*. Fungal Biol 2014;118(1):32–47.

[19] Li B, Yang Y, Cai J, Liu X, Shi T, Li C, et al. Genomic characteristics and comparative genomics analysis of two Chinese *Corynespora cassiicola* strains causing *Corynespora* leaf fall (CLF) disease. J Fungi (Basel) 2021;7(6):485. https://doi.org/10.3390/jof7060485.

[20] Munoz-Gomez SA, Wideman JG, Roger AJ, Slamovits CH. The origin of mitochondrial cristae from Alphaproteobacteria. Mol Biol Evol 2017;34(4):943–56.

[21] Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol 2003;29(3):380–95.

[22] Liu W, Cai Y, Zhang Q, Chen L, Shu F, et al. The mitochondrial genome of *Morchella importuna* (272.2 kb) is the largest among fungi and contains numerous introns, mitochondrial non-conserved open reading frames and repetitive sequences. Int J Biol Macromol 2020;143:373–81.

[23] Pramateftaki PV, Kouvelis VN, Lanaridis P, Typas MA. The mitochondrial genome of the wine yeast *Hanseniaspora uvarum*: a unique genome organization among yeast/fungal counterparts. FEMS Yeast Res 2006;6(1):77–90.

[24] Liu W, Cai Y, Zhang Q, Shu F, Chen L, Ma X, et al. Subchromosome-scale nuclear and complete mitochondrial genome characteristics of *Morchella crassipes*. Int J Mol Sci 2020;21(2):483.

[25] Salavirta H, Oksanen I, Kuuskeri J, Mäkelä M, Laine P, Paulin L, et al. Mitochondrial genome of *Phlebia radiata* is the second largest (156 kbp) among fungi and features signs of genome flexibility and recent recombination events. PLoS ONE 2014;9(5):e97141.

[26] Lang BF, Laforest M-J, Burger G. Mitochondrial introns: a critical view. Trends Genet 2007;23(3):119–25.

[27] Hafez M, Hausner G, Bonen L. Homing endonucleases: DNA scissors on a mission. Genome 2012;55(8):553–69.

[28] Wu P, Bao Z, Tu W, Li L, Xiong C, Jin X, et al. The mitogenomes of two saprophytic *Boletales* species (*Coniophora*) reveals intron dynamics and accumulation of plasmid-derived and non-conserved genes. Comput Struct Biotechnol J 2021;19:401–14.

[29] Li Q, Wu P, Li L, Feng H, Tu W, Bao Z, et al. The first eleven mitochondrial genomes from the ectomycorrhizal fungal genus (*Boletus*) reveal intron loss and gene rearrangement. Int J Biol Macromol 2021;172:560–72.

[30] Deng Y, Wu X, Wen D, Huang H, Chen Y, et al. Intraspecific mtDNA comparison of mycopathogen *Mycogone perniciosa* provides the insight into mitochondrial tRNA introns. Phytopathology 2021;111(4):639–48.

[31] Chen C, Li Q, Fu R, Wang J, Fan Z, Chen X, et al. Characterization of the complete mitochondrial genome of *Corynespora cassiicola* (Pleosporales: Dothideomycetes), with its phylogenetic analysis. Mitochondrial DNA B 2019;4(2):2938–9.

[32] Looi HK, Toh YF, Yew SM, Na SL, Tan YC, et al. Genomic insight into pathogenicity of dematiaceous fungus *Corynespora cassiicola*. PeerJ 2017;5:e2841.

[33] Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34(17):i884–90.

[34] Xu H, Luo X, Qian J, Pang X, Song J, Qian G, et al. FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS ONE 2012;7(12):e52249.

[35] Liu YC, Schroder J, Schmidt B. Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data. Bioinformatics 2013;29(3):308–15.

[36] 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.

[37] Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucl Acids Res 2013;41(13):e129.

[38] Valach M, Burger G, Gray MW, Lang BF. Widespread occurrence of organelle genome-encoded 5S rRNAs including permuted molecules. Nucl Acids Res 2014;42(22):13764–77.

[39] Donath A, Juhling F, Al-Arab M, Bernhart SH, Reinhardt F, et al. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucl Acids Res 2019;47(20):10543–10552.

[40] Araújo DS, De-Paula RB, Tomé LMR, Quintanilha-Peixoto G, Salvador-Montoya CA, Del-Bem L-E, et al. Comparative mitogenomics of Agaricomycetes: Diversity, abundance, impact and coding potential of putative open-reading frames. Mitochondrion 2021;58:1–13.

[41] Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol 2019;1962:1–14.

[42] Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW––a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res 2013;41(Web Server issue):W575–581.

[43] Lu J, Salzberg SL, Rzhetsky A. SkewIT: the skew index test for large-scale GC Skew analysis of bacterial genomes. PLoS Comput Biol 2020;16(12):e1008439.

[44] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger data sets. Mol Biol Evol 2016;33(7):1870–4.

[45] Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 2017;34(12):3299–3302.

[46] Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 2011;12:402.

[47] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and clustal X version 2.0. Bioinformatics 2007;23(21):2947–8.

[48] Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform 2019;20(4):1160–6.

[49] Vaidya G, Lohman DL, Meier R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 2011;27(2):171–80.

[50] Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol 2017;34(3):772–3.

[51] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30(9):1312–3.

[52] Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 2012;61(3):539–542.
[53] Franco MEE, López SMY, Medina R, Lucentini CG, Troncozo MI, Pastorino GN, et al. The mitochondrial genome of the plant-pathogenic fungus *Stemphylium lycopersici* uncovers a dynamic structure due to repetitive and mobile elements. PLoS ONE 2017;12(10):e0185545.
[54] Deng G, Zou Q, Chen Y, Wang L, Huang Ge, Cui Y, et al. The complete mitochondrial genome of *Cochliobolus miyabeanus* (Dothideomycetes, Pleosporaceae) causing brown spot disease of rice. Mitochondrial DNA B 2019;4(2):2832–3.
[55] Stone CL, Frederick RD, Tooley PW, Luster DG, Campos B, Winegar RA, et al. Annotation and analysis of the mitochondrial genome of *Coniothyrium glycines*, causal agent of red leaf blotch of soybean, reveals an abundance of homing endonucleases. PLoS ONE 2018;13(11):e0207062.
[56] Song N, Geng Y, Li X. The mitochondrial genome of the phytopathogenic fungus *Bipolaris sorokiniana* and the utility of mitochondrial genome to infer phylogeny of Dothideomycetes. Front Microbiol 2020;11:863.
[57] Shen X-Y, Li T, Chen S, Fan Li, Gao J, Hou C-L, et al. Characterization and phylogenetic analysis of the mitochondrial genome of *Shiraia bambusicola* reveals special features in the order of Pleosporales. PLoS ONE 2015;10(3):e0116466.
[58] Goodwin SB, McCorison CB, Cavaletto JR, Culley DE, LaButti K, Baker SE, et al. The mitochondrial genome of the ethanol-metabolizing, wine cellar mold *Zasmidium cellare* is the smallest for a filamentous ascomycete. Fungal Biol 2016;120(8):961–74.
[59] Eo J-K, Park E-S. The complete mitogenome of *Sydowia polyspora*. Mitochondrial DNA B 2019;4(1):1992–3.
[60] Gomes FEES, Arantes TD, Fernandes JAL, Ferreira LC, Romero H, Bosco SMG, et al. Polymorphism in mitochondrial group I introns among *Cryptococcus neoformans* and *Cryptococcus gattii* genotypes and its association with drug susceptibility. Front Microbiol 2018;9:86.
[61] Férandon C, Moukha S, Callac P, Benedetto J-P, Castroviejo M, Barroso G, et al. The *Agaricus bisporus cox1* gene: the longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. PLoS ONE 2010;5(11):e14048.
[62] Sumabat LG, Kemerait RC, Brewer MT. Phylogenetic diversity and host specialization of *Corynespora cassiicola* responsible for emerging target spot disease of cotton and other crops in the southeastern United States. Phytopathology 2018;108(7):892–901.
[63] Jin E-H, Sung JK, Lee S-I, Hong JH. Mitochondrial NADH dehydrogenase subunit 3 (*MTND3*) polymorphisms are associated with gastric cancer susceptibility. Int J Med Sci 2018;15(12):1329–33.
[64] Yang M, Zhang H, van der Lee TAJ, Waalwijk C, van Diepeningen AD, Feng J, et al. Population genomic analysis reveals a highly conserved mitochondrial genome in *Fusarium asiaticum*. Front Microbiol 2020;11:839.
[65] Ding Yu, Teng Y-S, Zhuo G-C, Xia B-H, Leng J-H. The mitochondrial tRNAHis G12192A mutation may modulate the clinical expression of deafness-associated tRNAThr G15927A mutation in a Chinese pedigree. Curr Mol Med 2019;19(2):136–46.