BMC Genomics

RESEARCH Open Access

Comparative analysis of mitochondrial genomes in the *Ceratocystidaceae* reveals highly conserved gene organization despite substantial genome size variation

Anien Viljoen[1], Tuan A. Duong[1], Aquillah M. Kanzi[1] and Brenda D. Wingfield[1*]

Abstract

Background Comparative studies of mitochondrial genomes can assist in determining evolutionary and phylogenetic relationships. In the *Ceratocystidaceae* fungal family, mitochondrial genomes from only two genera are publicly available. In this study, mitochondrial genomes from 10 genera in *Ceratocystidaceae* were characterised and comparatively analysed.

Results Mitochondrial (mt) genomes were fully assembled and characterized for 18 species and 10 genera of the *Ceratocystidaceae*. These mt genomes displayed a remarkable level of diversity owing to the variation in the intergenic regions and intron, both in size and number. Despite large variation in size, these mt genomes displayed high levels of synteny and sequence conservation in the oxidative phosphorylation (OXPHOS) genes and the ribosomal RNA and tRNA genes. The genes are found in identical order and orientation in all mt genomes characterized and the coding sequence of all the OXPHOS genes are highly conserved. Phylogenetic analysis of mitochondrial and nuclear DNA sequences reveals significant differences in evolutionary relationships among species and genera of *Ceratocystidaceae*.

Conclusion The remarkable variation in mt genome sizes in species of the *Ceratocystidaceae* is due to the variation observed in the intergenic regions and intron size and number in the conserved protein-coding genes. The gene order and content are identical across all species investigated.

Keywords *Ceratocystidaceae*, Synteny, Group I and II introns, Phylogeny, Genome size variation

*Correspondence:
Brenda D. Wingfield
brenda.wingfield@up.ac.za
[1]Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa

Background

The *Ceratocystidaceae* (Ascomycota) represents a large assemblage of plant and insect associated fungi. This family is currently divided into ten genera, all of which—except *Huntiella*—include plant pathogens [1–4]. The most notable genus in the family is *Ceratocystis*, which comprises highly aggressive pathogens of forestry and agriculturally important crops [1]. Noteworthy species include *Ceratocystis fimbriata*, which causes black rot of sweet potatoes [5]; *Ceratocystis platani*, the causal agent of canker and wilt of plane trees [6, 7]; *Ceratocystis manginecans*, which causes causing wilt disease in various hosts [8]; and *Ceratocystis lukuohia* which is responsible for large scale 'ōhi'a lehua mortality in in the Hawaiian Islands [9].

The mitochondrion essentially has the same functions in all eukaryotic cells [10]. It is responsible for ATP production and the translation of mitochondrial proteins [10–12]. In fungi, these processes are driven by 14 conserved protein coding genes, also referred to as the OXPHOS genes. These genes include ATP synthase subunit proteins *(atp6*, *atp8* and *atp9*), the cytochrome oxidase subunits (*cox1*, *cox2* and *cox3*), the cytochrome b gene (*cob*) and the seven NADH dehydrogenase subunits genes (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*) [10, 13, 14]. In addition to the OXPHOS genes, fungal mitochondrial genomes typically contain ribosomal protein S3 (RPS3), the large and small subunit of ribosomal RNA genes (*rns* and *rnl*, respectively) and a variable number of transfer RNA (tRNA) genes [10, 15].

Studies on mitochondrial (mt) genomes in fungi are fewer than those in animals and plants. Despite that, differences in mt genome sizes and variations have been relatively well documented across the three eukaryotic kingdoms [12]. Animal mt genomes are typically smaller, gene-dense and intron-poor when compared to fungal mitochondrial genomes [12], while plant mt genomes are larger, contain large intergenic regions and a highly variable number of introns [10, 16]. The genome size variation observed in fungal mt genomes is greater than that in plants and animals [10]. Fungal mt genome size variation has been reported between strains of the same species and between species of the same genus, although the latter tends to be more pronounced [10, 17].

The variability in the number and size of introns found in the conserved protein-coding genes highly contributes to the mitochondrial genome size variation observed in fungi [18–20]. Both group I and group II introns are found in fungi [10, 15] with group I introns being more common. This is in contrast with plant mitochondrial genomes where mainly group II introns are found [21]. Group I introns contain homing endonuclease genes (HEGs) which code for DNA-cleaving enzymes that are responsible for the movement of these introns [21]. In fungi, mitochondrial HEGs are classified as LAGLIDADG or GIY-YIG endonucleases. The naming corresponds to the conserved sequence motifs present in the endonuclease sequence [21, 22]. The presence of large intergenic regions is a common characteristic of mitochondrial genomes from plants and fungi despite fungi being more closely related to animals [12].

Despite the increasing availability of whole genome data for species of the *Ceratocystidaceae*, there are only a few mt genomes that have been fully assembled and characterised for these fungi. These include mt genome for *Endoconidiophora resinifera* [23] and those from several other *Ceratocystis* species [24, 25]. This limits the knowledge of the structure, organization and evolution of mitochondrial genomes across the family of *Ceratocystidaceae* species. Mitochondrial genomes are ideal for comparative studies, determining evolutionary relationship and resolving phylogenetic incongruences due to the increased evolutionary rate and the relatively small genome sizes [11, 26].

The aim of the study was thus to characterise and comparatively analyse mitochondrial genomes of species residing in 10 genera in the *Ceratocystidaceae*. Comparative analysis on the content, order and organisation of genes and genome size variation, more specifically the amount and type of introns and ORFs were conducted. Additionally, a phylogenetic relationship was inferred from 14 conserved mt genes and compared to the phylogeny constructed using a phylogenomic approach based on nuclear genes.

Results

Mitochondrial genome size, gene content and organization

Complete mitochondrial (mt) genomes were successfully assembled and annotated for 18 species. The mt genomes were assembled for 2 species per genera including *Ceratocystis*, *Berkeleyomyces*, *Chalarop*sis, *Endoconidiophora*, *Davidsoniella*, *Thielaviopsis*, *Huntiella*, and *Ambrosiella*, and one for *Catunica adiposa* and one for *Bretziella fagacearum*. The mt genomes from species of *Endoconidiophora* were the largest with assembled mt genome size of 225 205 bp and 215 106 bp for *E. laricicola* and *E. resinifera*, respectively. *Ceratocystis manginecans* has the smallest mt genome with 98 989 bp (Fig. 1). The mt genome GC contents were similar among the 18 species investigated with the highest being 31.38% for *H. abstrusa* and the lowest 26.42% for *Chalaropsis populi* (Table 1).

The conserved protein coding genes were all found on the same mtDNA strand as is typical in most ascomycetes [12]. Thirteen conserved fungal mt protein coding genes (*atp6*, *atp8*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*) were found in all 18

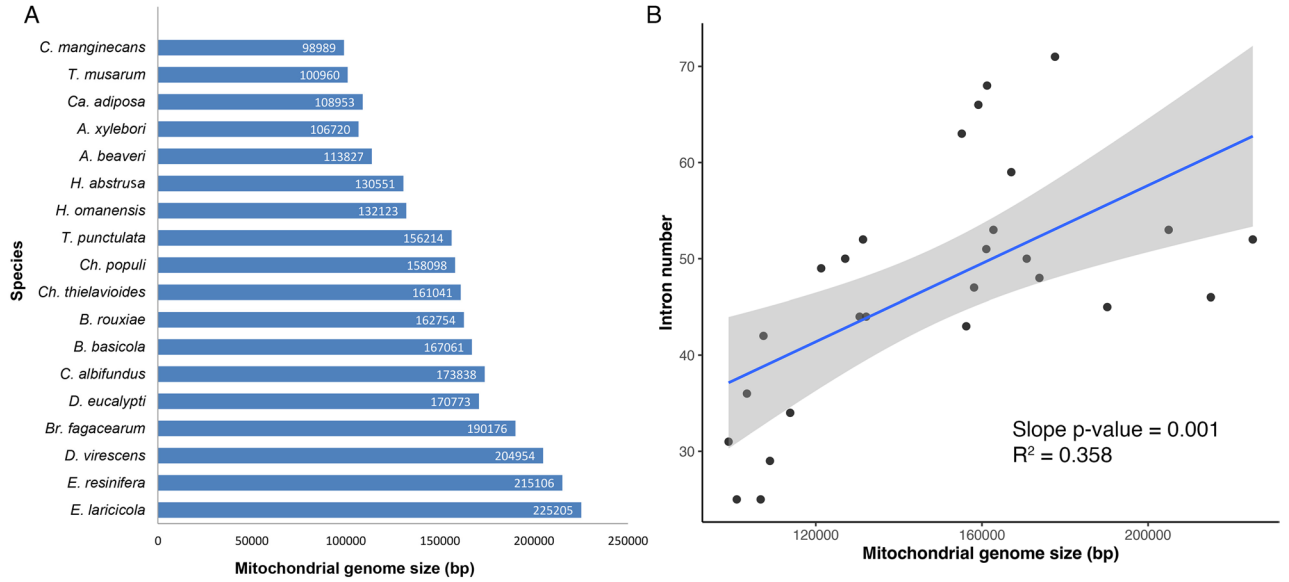

Fig. 1 The distribution of mitochondrial genome size of representative species of the *Ceratocystidaceae* (**A**). Correlation between the number of introns and mitochondrial genome size (**B**)

Table 1 Mitochondrial genome characteristics of species in the *Ceratocystidaceae* generated in this study. The genome sizes (bp = base pairs), the number of OXPHOS genes, tRNA, rRNA, ORFs found between the protein coding genes (intergenic ORFs), ORFs in the introns within the protein coding genes (intronic ORFs) and overlapping genes are provided

Species	GenBank accession	Size (bp)	GC (%)	Protein coding genes	tRNA genes	rRNA genes	Inter-genic ORFs	Intronic ORFs	Overlapping genes	
									NAD2-NAD3	NAD4L-NAD5
C. albifundus	PZ018566	173 838	27.79	14	28	2	19	59	Yes	Yes
C. manginecans	PZ018555	98 989	26.81	14	28	2	6	33	Yes	Yes
B. basicola	PZ018551	167 061	26.89	14	27	2	12	65	Yes	Yes
B. rouxiae	PZ018552	162 754	26.84	14	27	2	15	82	Yes	Yes
Ch. populi	PZ018550	158 098	26.42	14	26	2	24	61	Yes	Yes
Ch. thielavioides	PZ018549	161 041	26.44	14	26	2	20	67	Yes	Yes
E. resinifera	PZ018556	215 106	29.20	14	24	2	30	66	Yes	Yes
E. laricicola	PZ018557	225 205	28.50	14	24	2	22	78	Yes	Yes
D. eucalypti	PZ018554	170 773	27.18	14	25	2	21	51	Yes	Yes
D. virescens	PZ018553	204 954	27.68	14	24	2	34	80	Yes	Yes
T. musarum	PZ018558	100 960	27.17	14	25	2	9	31	Yes	Yes
T. punctulata	PZ018559	156 214	27.12	14	25	2	15	61	Yes	Yes
H. omanensis	PZ018561	132 123	30.21	13	25	2	8	23	Yes	No
H. abstrusa	PZ018560	130 551	31.38	13	25	2	4	21	Yes	No
A. *xylebori*	PZ018562	106 720	26.88	13	27	2	16	29	Yes	Yes
A. *beaveri*	PZ018563	113 827	26.85	13	25	2	11	37	Yes	Yes
Ca. adiposa	PZ018564	108 953	29.01	13	25	2	8	29	Yes	Yes
Br. fagacearum	PZ018565	190 176	30.28	14	26	2	34	63	Yes	Yes

species' mt genomes. The fourteenth conserved fungal mt gene, *atp9*, was found to be partial (only between 28 and 72 amino acids) in 13 out of the 18 species. Five species, namely *A. beaveri*, *A. xylebori*, *H. omanensis*, *H. abstrusa* and *Ca. adiposa*, were found to not have *apt9* in their mt genomes. The large (*rnl*) and small (*rns*) ribosomal subunit RNA genes were identified in all mt genomes of 18 species at a conserved genomic location but varying in sizes. The *rnl* genes are located between the *nad6* and *nad2* genes, while the *rns* genes are located between the *atp6* and *cox3* genes.

The order and direction of the conserved protein-coding and ribosomal RNA genes were conserved in all 18 mt genomes which is *rnl-nad2-nad3-atp9-cox2-nad4L-nad5-cob-cox1-nad1-nad4-atp8-atp6-rns-cox3-nad6*. The NADH dehydrogenase genes, *nad2-nad3* and *nad4L-nad5*, overlapped with a single nucleotide in majority

of the assembled mt genomes, except for *Huntiella* spp. where *nad4L* and *nad5* did not overlap.

Comparative analysis

tRNA clusters

All mt genomes contained tRNAs coding for all 20 essential amino acids, although the number of tRNAs varied between 24 and 28 (Fig. 2). The majority of the tRNAs are found in 3 clusters that are conserved across the species. These clusters contain more than 2 adjacent tRNAs and in some cases interrupted by the presence of intergenic open reading frames (ORF). This is especially evident in the largest cluster which is found between the *rnl* and *nad2* genes. This cluster contains up to 12 tRNA genes, namely *tRNA-Thr*, *tRNA-Glu*, *tRNA-Leu*, *tRNA-Met*, *tRNA-Leu*, *tRNA-Ala*, *tRNA-Phe*, *tRNA-Leu*, *tRNA-Leu*, *tRNA-Gln*, *tRNA-His*, *tRNA-Met* (Fig. 2). All 12 tRNA genes were present as in the cluster in the mt genomes of 4 out of the 18 species, namely *C. albifundus*, *C. manginecans*, *B. basicola* and *B. rouxiae*. For the remaining 14 species, the number of tRNA genes in this cluster range from 10 to 11. Intergenic ORFs that interrupt this cluster were found in 11 of the 18 species.

The second cluster is located between the *cox3* and *nad6* genes and is the most conserved. It contains 5 tRNA genes, *tRNA-Gly*, *tRNA-Lys*, *tRNA-Asp*, *tRNA-Ser*, *tRNA-Trp*. In *C. albifundus* and *Br. fagacearum*, a single intergenic ORF interrupts this cluster at the same position between tRNA-Asp and tRNA-Ser.

The third cluster shows little variation and is made up of up to 6 tRNA genes namely *tRNA-Val*, *tRNA-Trp*, *tRNA-Ile*, *tRNA-Ser*, *tRNA-Asn* and *tRNA-Pro* and is found between the *nad6* and *rnl* genes. Only *C. albifundus* and *C. manginecans* contain all 6 tRNA genes; the rest of the mt genomes have only 5 and are missing the *tRNA-Trp*. *Endoconidiophora resinifera* has an intergenic ORF in this region, which splits this cluster in two sub-clusters.

Conserved single tRNAs were found between specific pairs of protein-coding genes across all mitochondrial genomes analyzed: *tRNA-Arg* between *cox2-nad4L*, *tRNA-Cys* between *nad5-cob* and *tRNA–Arg* between

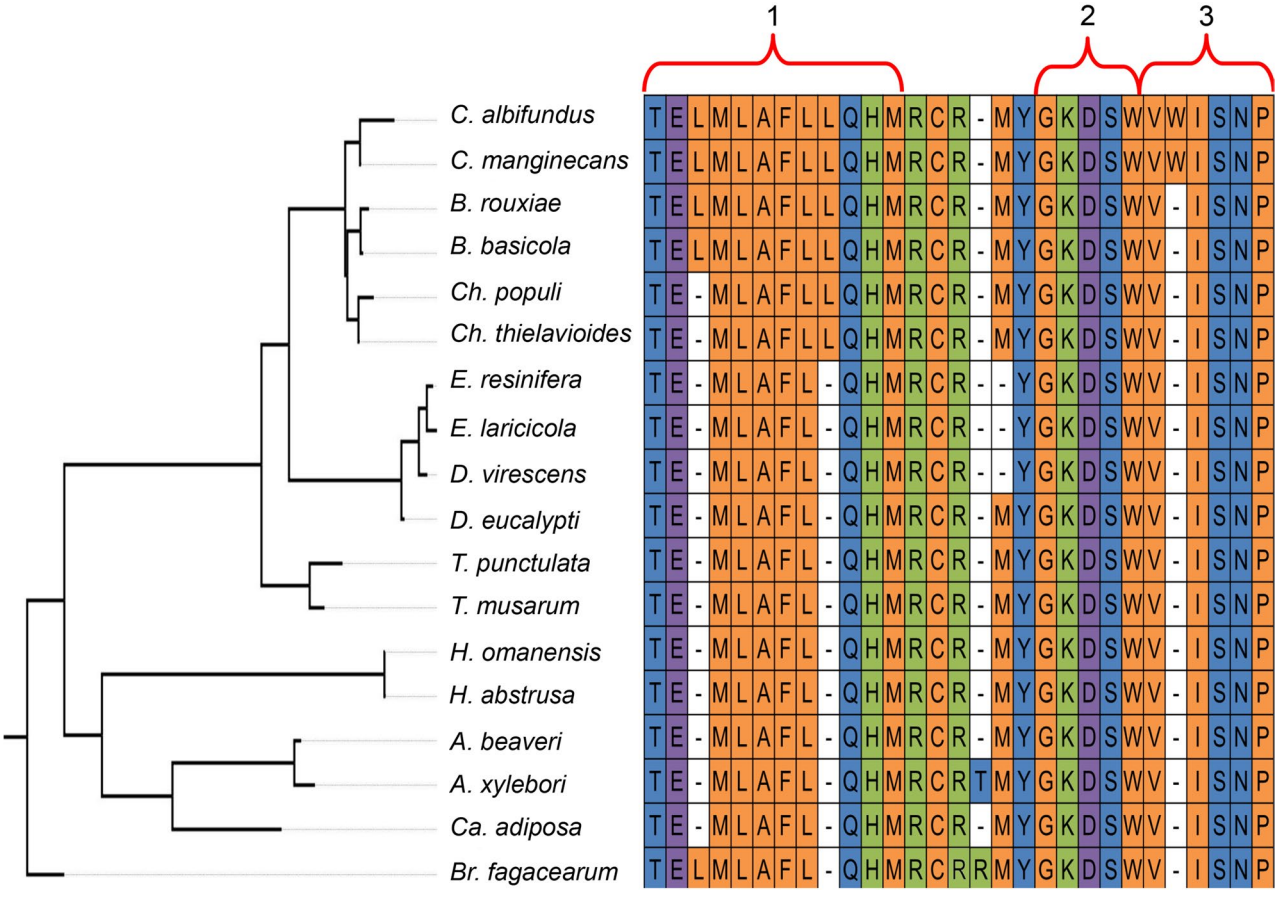

Fig. 2 The tRNA clusters found in the mitochondrial genomes of 18 species of the *Ceratocystidaceae*. The order of the tRNA genes is representative of the order in the mitochondrial genomes. The tRNA and the amino acid recognized are represented by the single letter standard abbreviation. The (-) indicates the absence of the particular tRNA. The red brackets indicate the three main tRNA clusters. The phylogenetic tree generated from mt gene sequences using IQ-TREE was used as backbone for this illustration

cox1-nad4. With the exception of the *Endoconidiophora* species and *D. virescens*, there are two tRNA genes, *tRNA-Met* and *tRNA-Tyr*, found between the *rns* and the *cox3* gene. The tRNA gene content and order is conserved, with the exceptions found within the tRNA clusters.

Introns, intergenic- and intron-encoded ORFs

The number of introns identified in the conserved protein coding genes involved in oxidative phosphorylation were highly variable. The smallest variation observed in number of introns is found in the ATP synthase encoding genes (*atp6*, *atp8* and *atp9*), which ranges from 0 to 3 introns per gene. The number of introns present in cytochrome b (*cob*) genes ranges from 2 to 11, whereas introns identified in the NADH dehydrogenase subunit genes (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5 and nad6*) ranges from 0 to 13 introns per gene. The highest variation in intron number was observed in the cytochrome c oxidase subunit genes (*cox1*, *cox2* and *cox3*) which range from 3 to 21 introns (Supplementary Table 1).

The introns were classified as Group IA (A or 5' partial), Group IB (5' partial, complete or 3' partial), Group IC, Group ID, Group I derived (A, B1, B2) or Group II (domain V). The majority of the species contained both Group I and II introns, with most of these belonging to Group I introns, which accounted for 86.43%, while Group II introns were significantly less abundant, accounting for only 2.69% of all introns. Several species, including *Ca. adiposa*, *T. musarum*, *T. punctulata*, *A. xylebori*, *Ch. populi* and *C. manginecans*, had only Group I introns. The percentage representation of each intron type was calculated for all 18 species and presented in Fig. 3.

The Group IB (complete) intron types are the most widespread intron type making up 33.94% of introns found in all 18 species, while Group IA (5' partial) is the rarest intron type with only 0.65%. The unknown intron types (10.38%) are mainly found in the *cox* and *nad* genes with the exceptions of *Br. fagacearum*, *D. eucalypti* and *A. beaveri* where the *cob* gene also contains unknown introns.

Open reading frames (ORFs) identified within the introns of the protein coding genes are called intronic ORFs; while ORFs found between the conserved genes are called intergenic ORFs. Intergenic ORFs are less common (24.76%) compared to intronic ORFs (75.24%). The number of ORFs identified in the *Ceratocystidaceae* species is highly variable as some introns can contain more than one ORF. Most of the identified ORFs in all 18 mt genomes were HEGs, LAGLIDADG or GIY-YIG family endonucleases and some were unknown. The unknown ORFs were labelled hypothetical protein and did not have any functional annotation that could be identified in any currently available protein database (Table 2). An ORF encoding for ribosomal protein S3 was found in the *rnl* gene in all 18 *Ceratocystidaceae* mt genomes. ORFs encoding for hypothetical proteins was mostly observed in two regions of the mt genomes. Firstly, between *rnl* and *nad2* gene in *B. rouxiae*, *D. virescens* and *Br. fagacearum* with an ORF size ranging between 99 and 118 amino acids. Secondly, between the *nad5* and *cob* genes in *E. resinifera*, *E. laricicola*, *D. virescens*, *T. musarum*, *T. punctulata* and *Br. fagacearum* ranging from 315 to 359 amino acids.

Codon usage

Codon usage and the frequency usage of each codon were determined for all mt genomes investigated. The

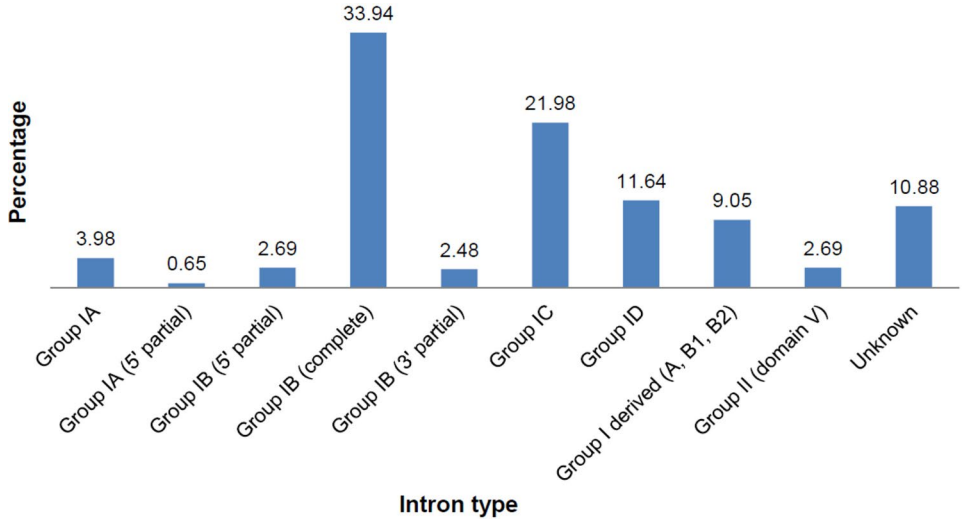

Fig. 3 Proportions of different intron types found in the mitochondrial genomes of 18 species

Table 2 ORFs encoding for hypothetical proteins. The species in which the hypothetical protein ORFs was found, ORF number, position in the mt genome (**intronic,** intergenic) and the size of the ORF. No homology or pfam domains could be identified for these ORFs on either NCBI or InterPro databases, respectively

Species	ORF	Genome position	Length of encoded protein
C. albifundus	ORF103	**Second intron of*nad2***	103
B. rouxiae	ORF114	**Third intron of*nad2***	114
″″	ORF99	**Fifth intron of*nad5***	99
″″	ORF99	Between *rnl* and *nad2*	99
Ch. thielavioides	ORF31	Between *rns* and *cox3*	31
E. resinifera	ORF317	Between *nad5* and *cob*	317
E. laricicola	ORF131	**Fifth intron of*nad5***	131
″″	ORF317	Between *nad5* and *cob*	317
D. virescens	ORF102	Between *rnl* and *nad2*	102
″″	ORF315	Between *nad5* and *cob*	315
″″	ORF101	Between *nad4* and *rns*	101
T. musarum	ORF317	Between *nad5* and *cob*	311
T. punctulata	ORF319	Between *nad5* and *cob*	312
Br. fagacearum	ORF160	**Seventh intron of*nad5***	160
″″	ORF118	Between *rnl* and *nad2*	118
″″	ORF105	Between *rnl* and *nad2*	105
″″	ORF359	Between *nad5* and *cob*	359

conserved protein coding genes in all mt genomes mostly favoured the start codon (ATG) and stop codon (TAA) with a few exceptions. The overlap between the *nad2* and *nad3* gene resulted in 10 out of 18 mt genomes having the *nad2* gene with TAG as the stop codon and subsequently resulting in the *nad3* gene having GTG as a start codon (Fig. 4). The TAG stop codon was used in the *cox2* gene in *T. punctulata*, in the *cob* gene for both the *Huntiella* species, in the *nad1* gene for *Ambrosiella* species, *Br. fagacearum* and *Ca. adiposa* and in *cox3* for *Br. fagacearum* and *Ca. adiposa*.

Conserved whole mitochondrial genome synteny

Comparative analysis of the conserved gene order of all 18 *Ceratocystidaceae* species' entire mt genomes showed a high level of end-to-end synteny (Fig. 5). The only variation observed was in the presence or absence of the *atp9* gene.

Phylogenetic analysis

Phylogenetic analyses using both supertree and supermatrix approaches resulted in identical phylogenies for each of the mitochondrial and nuclear datasets (Fig. 6). Several differences in evolutionary relationships were observed between nuclear phylogeny and mt phylogeny (Fig. 6). The most notable and well supported difference was in the placement of *Thielaviopsis* in which it was sister to the clade of *Ceratocystis*, *Chalaropsis*, *Berkeleyomyces*, *Endoconidiophora* and *Davidsoniella* in the mt phylogeny but it was sister to the clade of only *Ceratocystis*, *Chalaropsis*, *Berkeleyomyces* in the nuclear phylogeny. Another notable difference was in the placement of *Bretziella* where it was sister to the clade of *Ambrosiella*, *Catunica* and *Huntiella* in the mt phylogeny where it was only sister to the clade of *Ambrosiella* and *Catunica* in the nuclear phylogeny (Fig. 6). Other minor differences were the placement of *Chalaropsis* in relation to *Ceratocystis* and *Berkeleyomyces* in the two phylogenies, and the polyphyly of *Davidsoniella* in the mt phylogeny.

Discussion

Fungal mt genomes sizes are known to be highly variable, sometimes even within the genus, due to variations in intergenic regions, protein coding genes' introns size and number, and number of rRNA genes [10, 16, 18]. The comparison and characterization of the 18 assembled mt genomes from species residing in 10 genera in *Ceratocystidaceae* in this study indicate that even though the synteny of the protein coding genes are highly conserved, the variability in the non-coding regions contributes to the mt genome size variation. Mt genome sizes ranged from 98 989 bp (*C. manginecans*) to 225 205 bp (*E. laricicola*) which is not unexpected given the wide range of mt genome size variation observed in other fungal species [10]. The mt genome size variation is also observed between species from the same genus as in the case of *C. manginecans* (98 989 bp) and *C. albifundus* (173 838 bp). The GC content ranging from 26.42% in *Ch. populi* to 31.38% in *H. abstrusa* is also consistent with the overall AT-rich nature of fungal mt genomes [15]. This highlights that there was a strong correlation between the mt genome size and the number of introns as the largest mt genome assembled was *E. laricicola* (225 205 bp) which has 78 introns, while the smallest mt genome *C. manginecans* (98 989 bp) only has 33 introns.

Thirteen protein coding OXPHOS genes were consistently found in all of the *Ceratocystidaceae* fungal species mt genomes, with the *atp9* gene being partially present in 13 of the 18 mt genomes. The *atp9* gene is absent completely in *A. beaveri*, *A. xylebori*, *H. omanensis*, *H. abstrusa* and *Ca. adiposa.* The OXPHOS genes are essential mt genes responsible for oxidative phosphorylation [11, 13]. The core set of conserved protein-coding genes includes cytochrome oxidase subunits 1, 2 and 3; apocytochrome b (*cob*); NADH dehydrogenase subunits 1, 2, 3, 4, 4 L, 5 and 6; and ATP synthase subunits 6, 8 and 9 [15]. The order of rRNA and OXPHOS genes (*rnl-nad2-nad3-atp9-cox2-nad4L-nad5-cob-cox1-nad1-nad4-atp8-atp6-rns-cox3-nad6*) remained conserved across all 18 mt genomes and a single nucleotide overlap is present for *nad2* and *nad3* genes, as well as for *nad4L* and *nad5* genes, with the latter being absent in only the *Huntiella*

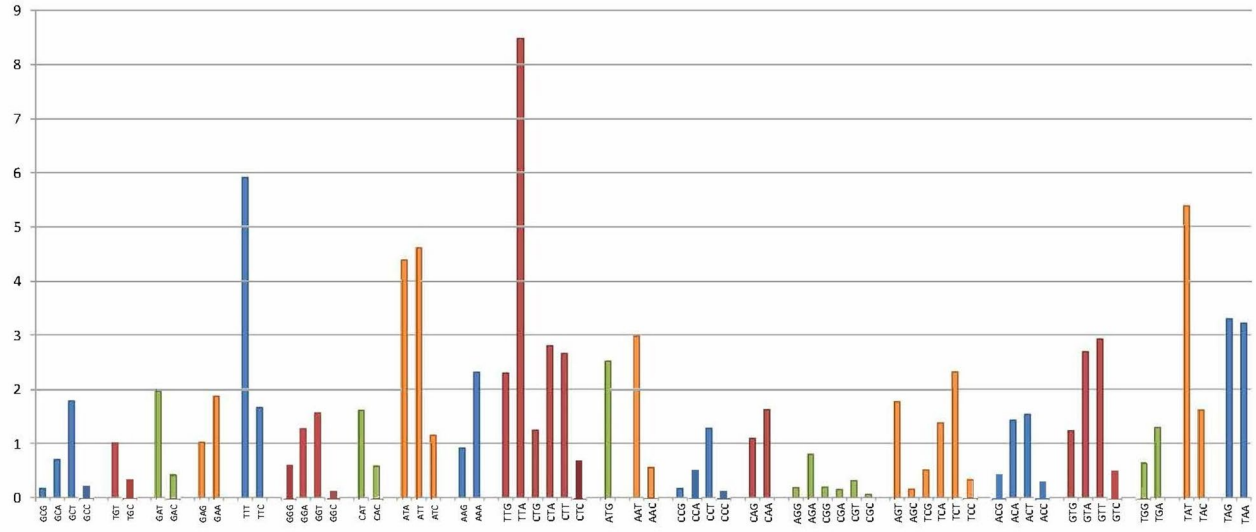

Fig. 4 Frequency of codon usage in OXPHOS genes in mitochondrial genomes of *Ceratocystidaceae* species. Codon frequency was calculated as a frequency per 1000 codons

species. The partial presence/absence of the *atp9* gene, which is widely conserved in fungal mt genomes, suggest possible gene loss or gene transference into the nuclear genome [27]. The *atp9* gene, usually involved in energy production, has been lost in several fungal species in the Ascomycota division, like the lichenized fungi found in the *Usnea* genus [28] and the pathogenic fungus *Stemphylium lycopersici*, where the *atp9* gene was found in the nuclear genome [29]. In most *Ceratocystidaceae* species, small sections of the *atp9* gene were found in both the mt genome as well as in the nuclear genome although neither genome contain a complete *atp9* gene.

Codon usage for the coding sequences of all the mt genomes was variable. The start codon ATG was overall favoured except in some cases where the *nad2* and *nad3* gene overlapped. Species of *Ceratocystis*, *Berkeleyomyces*, *Chalaropsis*, *Thielaviopsis*, *Br. fagacearum* and *Ca. adiposa* contained a single nucleotide overlap in which *nad2* has the TAG stop codon, resulting in GTG being used as start codon for *nad3* gene. While ATG is the start codon most frequently observed, the use of alternative start codons like GTG and TTG has been observed in other fungal species [30, 31]. The traditional mt stop codon (TAA) was observed in most of the coding genes except for a few genes in closely related species' mt genomes. The TAG stop codon was favoured in the *cox2* gene of *T. punctulata*, the *cob* gene in both *Huntiella* species' mt genomes, the *nad1* gene in both *Ambrosiella* species, the *nad1* and *nad6* in *Ca. adiposa* and lastly in the *nad1*, *nad3* and *nad6* genes in *Br. fagacearum*.

The synteny figure constructed from pairwise BLAST comparisons illustrates highly conserved and syntenic regions in all the mt regions especially for the OXPHOS genes in the *Ceratocystidaceae* and how the number of introns contributes to the mt genome size variability observed. The coding sequences of these OXPHOS genes remained relatively conserved in all 18 assembled mt genomes. When looking at the *cox1* there is a slight variation in the encoded protein size ranging from 527 to 537 amino acids but the number of introns found within this gene varies greatly between all the assembled mt genomes. The number of introns found in the *cox1* gene ranges from 9 in *C. manginecans* and *A. xylebori* to 21 in *E. laricicola.* This pattern of variation has been observed

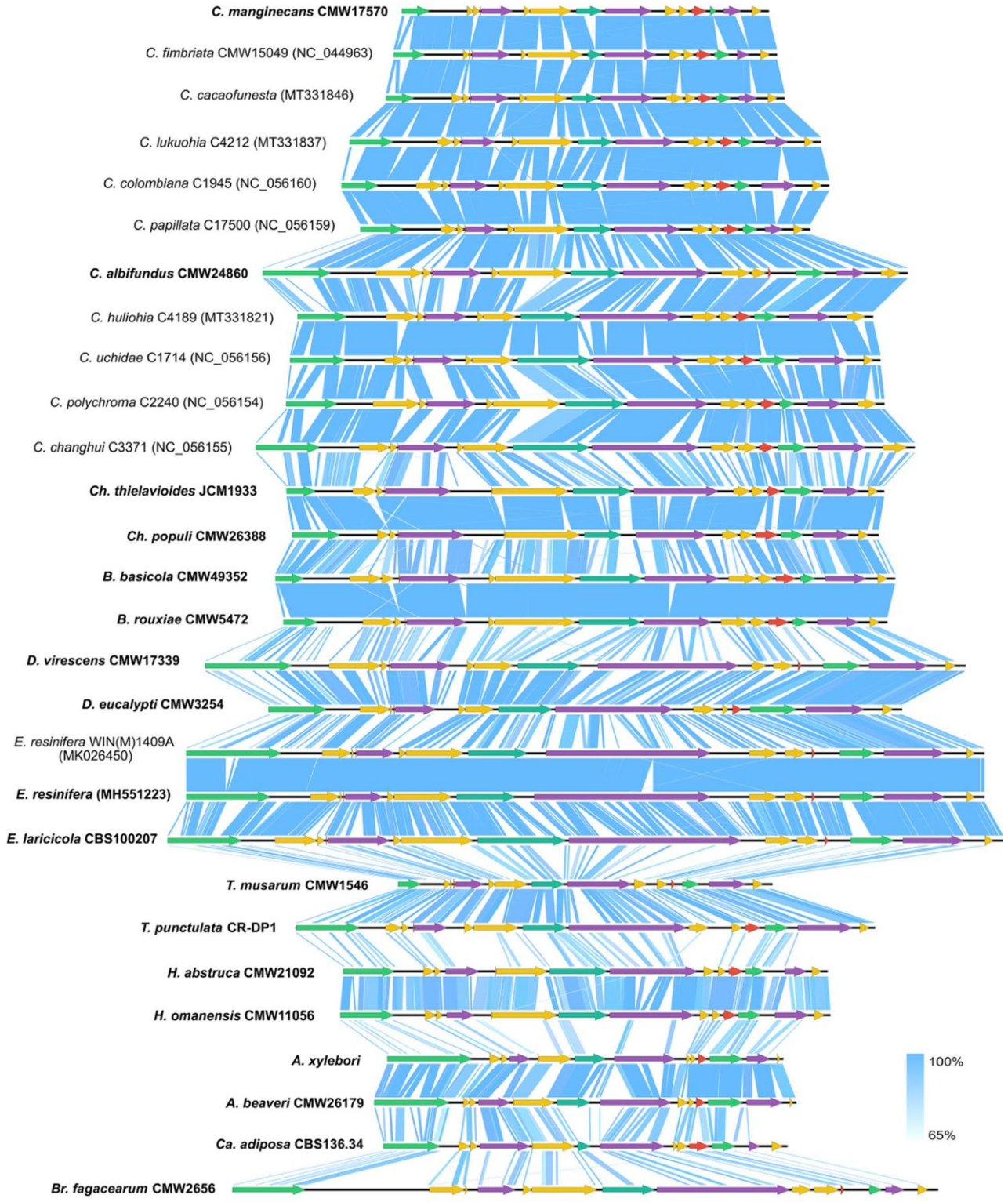

Fig. 5 Mitochondrial genome synteny of *Ceratocystidaceae* species. Green arrows indicate the *rnl* and *rns* genes, yellow arrows indicate *nad* genes, purple arrows indicate *cox* genes, red arrows indicate *atp* genes, and dark blue arrows indicate *cob* genes. Synteny blocks (ranging from 65% to 100%) identified with blastn are indicated by blue ribbons. Species with bold font type indicate mitochondrial genomes assembled and characterized in this study. Figure was generated using Easyfig

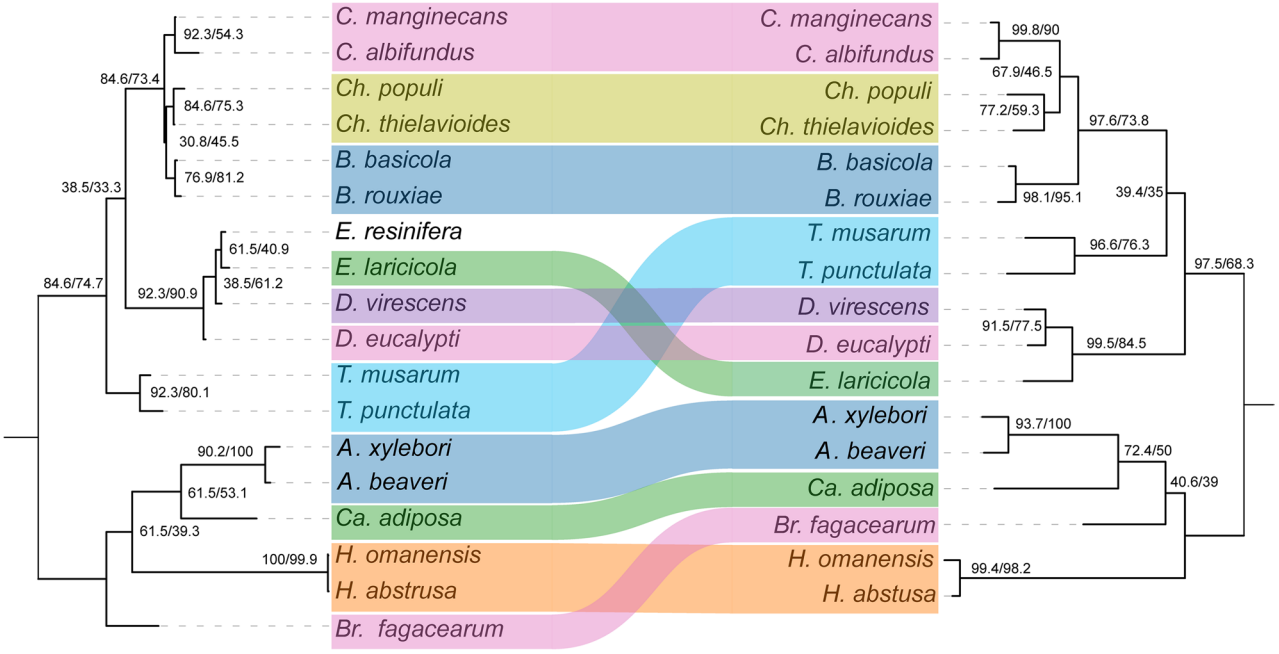

Fig. 6 Phylogenetic comparison between the mitochondrial phylogeny and nuclear phylogeny of *Ceratocystidaceae* species. The gene concordance factor (gCF) and site concordance factor (sCF) support values can be found at the nodes (gCF/sCF)

in all the other protein coding genes, with the exception of *atp8*, as it is the only OXPHOS gene that has no introns in any of the mt genomes investigated. The coding sequences of all the OXPHOS genes as well as the small (*rns*) and large (*rnl*) ribosomal subunit genes are conserved across all the genomes, the largest contributing factor to the genome size variation is the variability of the number of introns.

The types and frequencies of introns found in the *Ceratocystidaceae* is not unexpected as it has been found previously that Group I introns are the most commonly found in fungal mt genomes [21]. The Group I introns could be further divided into subgroups of which Group IB was found to be the most abundant making up a third of all introns present in all 18 mt genomes. Group II introns are rare in fungal mt genomes but have been detected in protein-coding genes and rRNA genes [32–34]. Group II introns in the 18 *Ceratocystidaceae* species investigated were identified in the *nad* and *cox* genes. More specifically, the *nad2* gene in *C. albifundus*, *B. basicola*, *B. rouxiae*, *Ch. thielaviodes*, *E. laricicola*, *D. eucalypti*, *D. virescens* and *Br. fagacearum*; as well as the *cox1* gene in *C. albifundus*, *E. resinifera*, *D. eucalypti*, *D. virescens*, *H. omanensis*, *H. abstrusa*, *A. beaveri* and *Br. fagacearum.* Additional Group II introns were found in the *nad6* gene of *C. albifundus* and in the *cox2* and *nad4* gene of *Br. fagacearum*.

Results of this study revealed significant differences in evolutionary relationships between genera of the *Ceratocystidaceae* when comparing mitochondrial- and nuclear genes phylogeny. Molecular sequence data can be used to clarify different evolutionary scenarios, and the choice of using genome data (such as mitochondrial or nuclear DNA) or functional sequences (like protein-coding versus non-coding genes) depends on the study's objective [35]. For the estimation of phylogenies of recently diverged species, mitochondrial gene sequences are a compelling option due to the rate of evolution and the linkage of genes as a single group, while nuclear genes evolve slower and genes can be selected from different chromosomes from which individual gene trees can be constructed making it a better option when deeper phylogenetic resolution is required [36, 37]. Several factors can contribute to incongruences between mitochondrial and nuclear gene tree phylogenies including horizontal gene transfer, gene duplication, introgressive hybridization, especially in recently diverged species and incomplete lineage sorting [38–41]. A study that compared the rate of substitution between a mitochondrial gene and its corresponding pseudogenes in the nuclear genome found that substitution in the mitochondrial gene can be up to 39 times faster, which can also contribute to discordance between mitochondrial and nuclear gene trees [42]. Mitochondrial DNA is almost exclusively inherited from the maternal parent and undergoes little to no recombination, but it has been shown that "leakage" of the paternal mitochondrial genome can happen during hybridization [43, 44]. There is thus the possibility of recombination between the maternal and paternal mitochondrial genomes, especially if there is cell

fusion between the parental species [38, 43, 44]. Horizontal transfer of genetic elements between mitochondrial genomes of different fungal species has also been observed, potentially contributing to discordant phylogenetic signals [24]. Phylogenetic analysis can also be further complicated by translocation of mitochondrial DNA into the nuclear genome, which will cause phylogenetic trees to be incongruent due to the high sequence similarity between the translocated mtDNA and organellar mtDNA [42].

Conclusion

The mitochondrial genomes of 18 species from 10 genera of the *Ceratocystidaceae* were successfully assembled and annotated in this study. The results indicate that while there was a large variation in genome size, there is high conservation of the conserved mitochondrial gene content and their order across all species investigated. The mitochondrial genome size variation observed is mostly due to the variation in the number and the sizes of the intron found in the intergenic regions and the conserved protein coding regions. Our results further support the link between mitochondrial genome sizes and the number of introns and how that plays a role in mitochondrial genome size variability. There are several factors that can thus contribute to incongruences in the topologies of phylogenetic trees constructed using mitochondrial DNA and nuclear DNA and the importance of scrutinizing both mitochondrial DNA and nuclear DNA is emphasized in order to better understand the phylogenetic relationship between closely related species and/or the evolutionary history of species.

Methods

Mt genome assembly

Raw sequence data for *Ceratocystis albifundus* CMW24860, *C. manginecans* CMW17570, *Davidsoniella eucalypti* CMW3254, *D. virescens* CMW17339, *Endoconidiophora laricicola* CBS100207, *Berkeleyomyces basicola* CMW49352, *Chalaropsis populi* CMW26388, *Bretziella fagacearum* CMW2656, *Catunica adiposa* CBS136.34, *Thielaviopsis musarum* CMW1546 and *Huntiella omanensis* CMW11056 were obtained from whole genome sequencing studies performed in Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria.

The whole genome data for *Chalaropsis thielavioides* JCM1933 (GenBank Accession number: GCA_001599435.1), *Thielaviopsis punctulata* CR-DP1 (GenBank Accession number: GCA_000968615.1), *Huntiella abstrusa* CMW21092 (GenBank Accession Number: GCA_025685495.1) and *Ambrosiella beaveri* CMW26179 (GenBank Accession Number: GCA_030294605.1) as well as the mitochondrial data from *Endoconidiophora resinifera* WIN(M)79 (GenBank Accession number: MH551223.1) were obtained from National Center for Biotechnology Information (NCBI).

All the above-mentioned raw sequence data are Illumina data, except for *Berkeleyomyces basicola* which is PacBio data, *B. rouxiae* which is Nanopore data and *Huntiella abstrusa* where a combination of PacBio and Illumina data was used to for the genome assembly.

Mitochondrial genome assembly using short read Illumina sequences

The Illumina raw read sequences were trimmed using Trimmomatic v0.32 [45] to remove all adaptor sequences and then assembled using SPAdes v3.14 [46]. Potential mt contigs were extracted based on the coverage. Contigs with overlapping flanking sequences were considered complete and those contigs were circularised and further analysed. Certain species' mt genomes could be assembled as one single contig using SPAdes whereas other species' mt genome assembly required additional steps.

The first additional strategy followed was using an iterative script that used the fragmented contigs from the initial SPAdes assembly as references. This assembly strategy has 3 steps where firstly the short read sequence data were mapped to the reference sequences using Bowtie v2.3.5.1 [47]. Secondly, SAMtools v1.9 [48] was used to extract the mapped reads and those with matching k-mers were identified using Mirabait. The last step which was contig extension and it involved the initial SPAdes assembly and the two above mentioned steps being repeated for 10 iterations.

For the assembly of mt genomes with unsuccessful de novo assembly, a combination of RaGOO [49] using species from the same genus as reference and GapFiller [50] was used.

Assembly using Oxford Nanopore sequences

The total genomic DNA from *B. rouxiae* (CMW5472) was extracted from freeze-dried mycelium using QIAGEN genomic tip as described in [51] with the exception that cellulase was omitted from the cell lysing step. Nanopore sequencing was carried out using the SQK-LSK109 library preparation kit (Oxford Nanopore) and sequenced on the FLO-MIN106 flow cell with the MinION Mk1B sequencer as described in Engelbrecht et al. (2021).

The low coverage genome Nanopore data for *B. rouxiae* was used to assemble the mt genome by using the de novo assembler Canu [52] followed by error correction using Racon [53] and two rounds of Medaka v 1.3.4 to create the consensus sequence. Potential mt contigs were identified based on expected higher sequence coverage compared to the nuclear contigs.

Assembly using PacBio sequences

The draft genome assembly of *Huntiella abstrusa* CMW21092 available on NCBI was used for mt assembly. The mt genome was extracted based on coverage and polishing with Pilon v1.23 [54] using the Illumina short reads available under the same accession number as the draft genome assembly to improve the quality of the mt genome.

Mt genome annotation

MFannot (https://megasun.bch.umontreal.ca/apps/mfannot/) was used to predict conserved genes, tRNA, large and small subunit RNA and introns using the genetic code 4 (Mold mitochondrial DNA). The putatively annotated mt genomes were manually inspected and curated on CLC Main Workbench v 8.1 using BLASTp against the non-redundant protein sequence database of the NCBI. For manual curation, the Sequence Manipulation Suite (https://www.bioinformatics.org/sms2/translate.html) translation function was used to translate DNA sequences to amino acid sequences.

Comparative genome analysis

All 18 assembled and annotated mt genomes from 10 different genera were comparatively analysed. The GC content for all 18 mt genomes was calculated using the GC content calculator (https://jamiemcgowan.ie/bioinf/gc.html). The tRNAs predicted by MFannot were comp ared between all 18 species. Introns were characterised as Group IA (A or 5' partial), Group IB (5' partial, complete or 3' partial), Group IC, Group ID, Group I derived (A, B1, B2) or Group II (domain V) by RNAweasel (https://megasun.bch.umontreal.ca/apps/rnaweasel/). Assig nment of intergenic and intronic ORFs as LAGLIDADG or GIY-YIG homing endonucleases or hypothetical proteins was done using BLASTp on NCBI (https://www.ncbi.nlm.nih.gov/) and InterPro (https://www.ebi.ac.uk/interpro/). Codon usages for the codon sequence of core set of conserved protein-coding genes of all 18 mt genomes were calculated using the codon function of the Sequence Manipulation Suite (https://www.bioinformatics.org/sms2/codon_usage.html). Synteny between the mt genomes were investigated by generating BLAST output file using BLASTn and visualising the relationship using Easyfig [55].

Sequence and phylogenetic analysis

Two different approaches were used to construct the phylogenetic trees for both the conserved genes from the mitochondria as well as the shared nuclear BUSCO genes. The phylogenetic trees were constructed using the supermatrix and supertree approach. The nuclear genomes of 17 publically available *Ceratocystidaceae* species were tested for completeness with BUSCO v4.0.5 using ascomycota_odb9 dataset [56]. There is no publicly available *Endoconidiophora resinifera* genome, thus excluding it from the phylogenetic analysis. The shared BUSCO genes that were present in 5 or more taxa were extracted. Both the coding sequences of the conserved mitochondrial genes and shared conserved nuclear genes were aligned using MAFFT [57] and trimmed with trimAl [58]. The trimmed and aligned data was used for both the supertree and supermatrix approaches. Individual gene trees were created for the supertree approach using IQ-TREE [59] and the individual trees combined into a single final phylogenetic tree using ASTRAL [60] for both the conserved mitochondrial genes and conserved nuclear genes respectively. Using IQ-TREE, the branch length estimation, gene concordance factor (gCF) and site concordance factor (sCF) [61, 62] was determined for both the resulting phylogenies.

The supermatrix approach used the aligned and trimmed data to create a data matrix using FASconCAT [63]. The best model of evolution was determined using PartitionFinder 2 [64], after which IQ-TREE was used to generate the phylogenetic trees for the conserved mt genes and the shared nuclear genes.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12755-2.

Supplementary Material 1. Supplementary Table 1. Gene size, coding sequence (CDS) size and the number of introns present in each of the 14 conserved protein-coding genes of all 18 mitochondrial genomes. Nd = not detectable.

Acknowledgements

Not applicable.

Authors' contributions

A.V. wrote the first draft of the manuscript. A.V., T.A.D and A.K did the analysis and figure preparation. BDW provided the funding. All authors contributed equally to the ideas behind the study and manuscript review.

Funding

Funding for this study was from the DSI-NRF SARChI chair in Fungal Genomics (National Research Foundation, SARCI150216114244).

Data availability

Mitogenomes of *Chalaropsis thielavioides* JCM1933 (GenBank accession number: GCA_001599435.1), *Thielaviopsis punctulata* CR-DP1 (GenBank accession number: GCA_000968615.1), *Huntiella abstrusa* CMW21092 (GenBank accession number: GCA_025685495.1) and *Ambrosiella beaveri* CMW26179 (GenBank Accession number: GCA_030294605.1) were obtained from National Center for Biotechnology Information (NCBI). Mitogenome of 18 species generated in this study were deposited to NCBI database and accession numbers are provided in Table 1.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication
All authors consent to the publication.

Competing interests
The authors declare no competing interests.

Received: 25 November 2024 / Accepted: 12 March 2026
Published online: 19 March 2026

References

1. De Beer ZW, Duong TA, Barnes I, Wingfield BD, Wingfield MJ. Redefining *Ceratocystis* and allied genera. Stud Mycol. 2014;79:187–219.
2. Nel WJ, Duong TA, Wingfield BD, Wingfield MJ, de Beer ZW. A new genus and species for the globally important, multihost root pathogen *Thielaviopsis basicola*. Plant Pathol. 2018;67:871–82.
3. Mayers CG, Harrington TC, Masuya H, Jordal BH, McNew DL, Shih H-H, et al. Patterns of coevolution between ambrosia beetle mycangia and the *Ceratocystidaceae*, with five new fungal genera and seven new species. Persoonia. 2020;44:41–66.
4. De Beer ZW, Marincowitz S, Duong TA, Wingfield MJ. Bretziella, a new genus to accommodate the oak wilt fungus, Ceratocystis fagacearum (Microascales, Ascomycota). MycoKeys. 2017;27:1–19.
5. Halsted B.D. Some fungus diseases of the sweet potato. New Jersey Agricultural College Experiment Station. 1890;76:25–7
6. Walter JM. Canker stain of plane trees. USDA Circular. Washington, DC: US Department of Agriculture; 1946;742.
7. Walter JM, Rex EG, Schreiber R. The rate of progress and destructiveness of canker stain of plane trees. Phytopathology. 1952;42:236–9.
8. Harrington TC, Ferreira MA, Somasekhara YM, Vickery J, Mayers CG. An expanded concept of *Ceratocystis manginecans* and five new species in the Latin American clade of *Ceratocystis*. Mycologia. 2024;116:184–212.
9. Barnes I, Fourie A, Wingfield MJ, Harrington TC, McNew DL, Sugiyama LS, et al. New *Ceratocystis* species associated with rapid death of *Metrosideros polymorpha* in Hawai'i. Persoonia. 2018;40:154–81.
10. Sandor S, Zhang Y, Xu J. Fungal mitochondrial genomes and genetic polymorphisms. Appl Microbiol Biotechnol. 2018;102:9433–48.
11. Burger G, Gray MW, Lang BF. Mitochondrial genomes: anything goes. Trends Genet. 2003;19:709–16.
12. Aguileta G, de Vienne DM, Ross ON, Hood ME, Giraud T, Petit E, et al. High variability of mitochondrial gene order among fungi. Genome Biol Evol. 2014;6:451–65.
13. Gray MW, Burger G, Lang BF. The origin and early evolution of mitochondria. Genome Biol. 2001;2:REVIEWS1018.
14. 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:e97141.
15. Hausner G. Fungal Mitochondrial Genomes, Plasmids and Introns. In: AroraDK, Khachatourians GG, editors. Applied Mycology and Biotechnology:Fungal Genomics, vol. III. New York: Elsevier Science; 2003. p. 101–31.
16. Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proc Natl Acad Sci U S A. 2000;97:6960–6.
17. Bullerwell CE, Lang BF. Fungal evolution: the case of the vanishing mitochondrion. Curr Opin Microbiol. 2005;8:362–9.
18. Friedrich A, Jung PP, Hou J, Neuvéglise C, Schacherer J. Comparative mitochondrial genomics within and among yeast species of the *Lachancea* genus. PLoS ONE. 2012;7:e47834.
19. Jung PP, Friedrich A, Reisser C, Hou J, Schacherer J. Mitochondrial genome evolution in a single protoploid yeast species. G3 (Bethesda). 2012;2:1103–11.
20. Zhang Y, Zhang S, Zhang G, Liu X, Wang C, Xu J. Comparison of mitochondrial genomes provides insights into intron dynamics and evolution in the caterpillar fungus *Cordyceps militaris*. Fungal Genet Biol. 2015;77:95–107.
21. Lang BF, Laforest M-J, Burger G. Mitochondrial introns: a critical view. Trends Genet. 2007;23:119–25.
22. 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:e0156104.
23. Zubaer A, Wai A, Hausner G. The mitochondrial genome of *Endoconidiophora resinifera* is intron rich. Sci Rep. 2018;8:17591.
24. Mayers CG, Harrington TC, Wai A, Hausner G. Recent and ongoing horizontal transfer of mitochondrial introns between two fungal tree pathogens. Front Microbiol. 2021;12:656609.
25. Syazwan SA, Mohd-Farid A, Yih Lee S, Mohamed R. Comparative analysis of mitochondrial genomes in *Ceratocystis fimbriata* complex across diverse hosts. Gene. 2024;921:148539.
26. Fourie G, van der Merwe NA, Wingfield BD, Bogale M, Tudzynski B, Wingfield MJ, et al. Evidence for inter-specific recombination among the mitochondrial genomes of *Fusarium* species in the *Gibberella fujikuroi* complex. BMC Genomics. 2013;14:605.
27. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 2003;29:380–95.
28. Funk ER, Adams AN, Spotten SM, Van Hove RA, Whittington KT, Keepers KG, et al. The complete mitochondrial genomes of five lichenized fungi in the genus Usnea (Ascomycota: Parmeliaceae). Mitochondrial DNA B Resour. 2018;3:305–8.
29. 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:e0185545.
30. Zhang Y-J, Zhang H-Y, Liu X-Z, Zhang S. Mitochondrial genome of the nematode endoparasitic fungus *Hirsutella vermicola* reveals a high level of synteny in the family *Ophiocordycipitaceae*. Appl Microbiol Biotechnol. 2017;101:3295–304.
31. Xiong C, Lin Y, Keyhani NO, Shang J, Mao Y, Yang J et al. Mitochondrial genomes from fungal the entomopathogenic Moelleriella genus reveals evolutionary history, intron dynamics and phylogeny. J Fungi. 2025;11(2):94.
32. Toor N, Zimmerly S. Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. RNA. 2002;8:1373–7.
33. Lambowitz AM, Zimmerly S. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb Perspect Biol. 2011;3:a003616.
34. Gonzalez P, Barroso G, Labarère J. Molecular gene organisation and secondary structure of the mitochondrial large subunit ribosomal RNA from the cultivated Basidiomycota *Agrocybe aegerita*: a 13 kb gene possessing six unusual nucleotide extensions and eight introns. Nucleic Acids Res. 1999;27:1754–61.
35. Springer MS, DeBry RW, Douady C, Amrine HM, Madsen O, de Jong WW, et al. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. Mol Biol Evol. 2001;18:132–43.
36. Moore WS. Infering phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution. 1995;49:718–26.
37. Lin C-P, Danforth BN. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol Phylogenet Evol. 2004;30:686–702.
38. Hogan RI, Hopkins K, Wheeler AJ, Yesson C, Allcock AL. Evolution of mitochondrial and nuclear genomes in *Pennatulacea*. Mol Phylogenet Evol. 2023;178:107630.
39. Ballard JWO, Whitlock MC. The incomplete natural history of mitochondria. Mol Ecol. 2004;13:729–44.
40. Shaw KL. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proc Natl Acad Sci U S A. 2002;99:16122–7.
41. Rubinoff D, Holland BS. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst Biol. 2005;54:952–61.
42. Arctander P. Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. Proc Biol Sci. 1995;262:13–9.
43. Whittaker SL, Assinder SJ, Shaw DS. Inheritance of streptomycin and chloramphenicol resistance in *Phytophthora infestans*: evidence for co-segregation of mitochondrial DNA and streptomycin resistance. Mycol Res. 1996;100:87–92.
44. Kroon LP, Bakker FT, Van Den Bosch GB, Bonants PJ, Flier WG, Flier WG. Phylogenetic analysis of *Phytophthora* species based on mitochondrial and nuclear DNA sequences. Fungal Genet Biol. 2004;41:766–82.
45. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
46. 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;19:455–77.
47. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
49. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 2019;20:224.
50. Nadalin F, Vezzi F, Policriti A. GapFiller: a de novo assembly approach to fill the gap within paired reads. BMC Bioinformatics. 2012;13(Suppl 14):S8.
51. Engelbrecht J, Duong TA, Prabhu SA, Seedat M, van den Berg N. Genome of the destructive oomycete *Phytophthora cinnamomi* provides insights into its pathogenicity and adaptive potential. BMC Genomics. 2021;22:302.
52. 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;27:722–36.
53. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017;27:737–46.
54. 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:e112963.
55. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27:1009–10.
56. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
57. Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–66.
58. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.
59. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.
60. Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics. 2014;30:i541–8.
61. Minh BQ, Hahn MW, Lanfear R. New methods to calculate concordance factors for phylogenomic datasets. Mol Biol Evol. 2020;37:2727–33.
62. Mo YK, Lanfear R, Hahn MW, Minh BQ. Updated site concordance factors minimize effects of homoplasy and taxon sampling.Updated site concordance factors minimize effects of homoplasy and taxon sampling. Bioinformatics. 2023;39:btac741.
63. Kück P, Meusemann K, FASconCAT. Convenient handling of data matrices. Mol Phylogenet Evol. 2010;56:1115–8.
64. 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;34:772–3.

Publisher's note