

SCIENTIFIC REPORTS



OPEN

The mitochondrial genome of *Endoconidiophora resinifera* is intron rich

Abdullah Zubaer, Alvan Wai & Georg Hausner

Endoconidiophora resinifera (= *Ceratocystis resinifera*) is a blue-stain fungus that occurs on conifers. The data showed that the *Endoconidiophora resinifera* mitochondrial genome is one of the largest mitochondrial genomes (>220 kb) so far reported among members of the Ascomycota. An exceptional large number of introns (81) were noted and differences among the four strains were restricted to minor variations in intron numbers and a few indels and single nucleotide polymorphisms. The major differences among the four strains examined are due to size polymorphisms generated by the absence or presence of mitochondrial introns. Also, these mitochondrial genomes encode the largest cytochrome oxidase subunit 1 gene (47.5 kb) reported so far among the fungi. The large size for this gene again can be attributed to the large number of intron insertions. This study reports the first mitochondrial genome for the genus *Endoconidiophora*, previously members of this genus were assigned to *Ceratocystis*. The latter genus has recently undergone extensive taxonomic revisions and the mitochondrial genome might provide loci that could be applied as molecular markers assisting in the identification of taxa within this group of economically important fungi. The large mitochondrial genome also may provide some insight on mechanisms that can lead to mitochondrial genome expansion.

Endoconidiophora resinifera (= *Ceratocystis resinifera*) is a fungus that belongs to the Ceratocystidaceae (Sordariomycetes, Microascales). It is associated with causing blue-stain on sapwood that ultimately leads to the discolouration of timber and timber-derived products. Blue-stain is considered one of the major causes of loss in value of conifer timber as it restricts its export potential and discoloured timber is less desirable for high end use¹. Some members of the genus *Ceratocystis sensu lato* (recently subdivided into several new genera including *Endoconidiophora*²) are known for causing infections such as black rot disease in sweet potato (*Ceratocystis fimbriata*)³, oak wilt (*Ceratocystis fagacearum*)⁴, wilt in cacao plant (*Ceratocystis cacaofunesta*)⁵, canker stain of plane trees (*Ceratocystis platani*)⁶, and sapstreak in maple tree (*Ceratocystis virescens*)⁷. *Endoconidiophora resinifera* has not been associated with any pathogenicity but this insect vectored fungus can colonize bark beetle galleries and wounds in species of *Picea*⁸. Species of *Ceratocystis s.l.* have been studied with regards to their taxonomy, blue-staining ability, and pathology⁹, but so far only *C. cacaofunesta* has been examined in more detail with regards to genomic investigations¹⁰. Previous studies on this group of fungi with regards to mitochondrial DNA focused on the rRNA genes and these displayed a large variety of intron insertions among various *Ceratocystis s.l.* species^{11,12}. Additional mitochondrial genomes have recently been sequenced for members of *Ceratocystis* but so far a detailed annotation is only available for the mitochondrial genome of *C. cacaofunesta*^{10,13–16}.

Fungal mitochondrial genomes usually encode genes involved in translation [small and large ribosomal subunit RNAs (*rns* and *rnl*) and tRNAs], proteins involved in the respiratory chain [subunits for Complex III and Complex IV (*cob*, *cox1*, *cox2*, and *cox3*)], subunits of NADH dehydrogenase (*nad1* to *nad6* and *nad4L*; except for members of the Taphrinomycota and some members of the Saccharomycetales), plus some of the components of the ATP synthase (*atp6*, *atp8*, and usually *atp9*), and in some instances the ribosomal protein RPS3^{17,18}. Mitochondrial genome sizes among the fungi are quite variable ranging from 12.055 kb (in *Rozella allomyces*)¹⁹ to 235 kb (in *Rhizoctonia solani*)²⁰. Mitochondrial genome size variation has also been reported among closely related species²¹. The size variations are mainly due to the number and sizes of intron insertions and size of intergenic spacers. Gene order, repeats, and in some instances other types of elements such as plasmid insertions are additional sources that generate variability among fungal mitochondrial genomes²².

Department of Microbiology, University of Manitoba, Winnipeg, MB, R3T 2N2, Canada. Correspondence and requests for materials should be addressed to G.H. (email: hausnerg@cc.umanitoba.ca)

Fungal mitochondrial introns, based on structure and splicing mechanisms, can be assigned to either group I or group II introns. These elements are potential ribozymes that can in part catalyze their own removal from transcripts; in addition these introns can encode open reading frames (ORFs) for so-called intron-encoded proteins (IEPs). Fungal mitochondrial group I introns tend to encode GIY-YIG or LAGLIDADG homing endonuclease genes (HEGs), and group II introns typically encode reverse transcriptase (RT) genes²³. These IEPs tend to catalyze the mobility of their respective introns from an intron-plus to an intron-minus cognate allele. Some IEPs have been shown to assist in the splicing of the introns that encode them by providing so called maturase activity; i.e. these IEPs promote the folding of the intron RNA into a splicing competent structure^{24,25}. Various nuclear genome-encoded factors also have been co-opted to assist in the splicing of mitochondrial introns²⁶.

Group I introns primarily mobilize via a DNA-based mechanism that involves its IEP [homing endonucleases (HEases)] generating a double-stranded cut at a cognate allele that is repaired by the double-strand break repair system. This involves homologous recombination using the intron-plus allele as the repair template and in a nonreciprocal manner the intron sequence and sometimes some flanking markers are transferred to repair the double-stranded break. Group II introns primarily act like retroelements where the mobility pathway utilizes an RNA intermediate and reverse transcriptase activity. In general, mobility of these group I and group II introns is referred to as homing or retrohoming, respectively, as they tend to invade cognate alleles that have not yet been invaded at a particular site. However, these elements can potentially insert into new locations (ectopic integration) here the terms transposition or retro-transposition (for group II introns) are applicable¹⁷. Mobile introns and homing endonuclease genes are sometimes referred to as selfish DNAs or selfish genes as they do not appear to benefit the genomes that host these elements. It is generally assumed that mobile introns are neutral with regards to phenotype thus ensuring their survival²⁶. However, being “neutral” (i.e. evolving by drift), can result in the rapid degeneration of these elements due to a lack of selection. In order for long term persistence, these elements have to invade intron-less alleles or invade new sites²⁷. There is also considerable evidence that these elements move horizontally across species barriers ensuring their long term persistence within populations or among fungal mitochondrial genomes²⁸.

The current study characterized the mitochondrial genomes of four strains of *E. resinifera*. The most noteworthy findings for these large (>214 kbp) genomes were the large numbers of introns and their IEPs. In addition, other components such as the protein-coding genes and other genetic components such as intergenic regions and remnants of inserted plasmids are also being described. These intron-rich genomes provide an opportunity to examine the mobility of group I introns and their HEGs. Currently, the actual mechanisms for intron acquisition and loss are still poorly understood. With regards to intron content one could speculate that ancestral mtDNAs were intron-rich and they are gradually being eroded and lost or alternatively, introns are continuously lost and reacquired by outcrossing or horizontal transfer^{27,28}. The *E. resinifera* mitochondrial genomes were compared among the four strains examined, and also with three related mitochondrial genomes from *C. platani*, *C. cacaofunesta* and *C. fimbriata* in order to gain a better understanding of the evolutionary mechanisms that could promote intron loss or gain and mitochondrial genome rearrangements.

Results

Mitochondrial genome of *E. resinifera*. The mitochondrial genomes for four *E. resinifera* strains [WIN(M)79 (=UAMH 9644), WIN(M)1409 A, WIN(M)1410B, WIN(M)1411] were sequenced, assembled and annotated (GenBank accession numbers: MH551223, MK026450, MK026449, MK012641 respectively). The mitochondrial genome for strain WIN(M)79 was annotated first and this genome was used as the reference mtDNA for this study; overall, the four *E. resinifera* mtDNAs only differed by the absence or presence of 4 introns and 2 single nucleotide polymorphisms. The genome of *E. resinifera* is composed of protein-coding genes such as (*atp6*, *atp8*, *cob*, *cox1-3*, *nad1-6* and *nad4L*) and rRNA (*rns* and *rnl*) and tRNA (27 tRNAs) genes. Most of the protein and rRNA -coding genes were noted to be populated with group I and group II introns, and most introns encode open reading frames (iORFs). The mitochondrial genomes of the *E. resinifera* strains were annotated in Artemis and the genome for strain WIN(M)79 was visualized in Circos (Fig. 1).

Protein-coding, rRNA and tRNA genes. The mitochondrial genome of *E. resinifera* contains 14 protein-coding genes and this includes NADH dehydrogenase subunits (*nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*) which contribute towards the electron transport complex I, cytochrome oxidase subunits (*cob*, *cox1*, *cox2* and *cox3*) that are part of complex III and complex IV, ATP synthase subunits (*atp6* and *atp8*) and the gene that encodes the 40 S ribosomal protein S3 (*rps3*). The genome contains the small and large ribosomal RNA genes (*rns* and *rnl*). A total of 27 tRNA genes were identified of which most are located around the *rnl* gene; 10 tRNA genes are upstream and 12 are downstream of the *rnl* gene although the upstream tRNA cluster is interrupted by the *nad6* gene. The remaining 5 tRNA genes are dispersed along the genome. All of the protein-coding genes, rRNA and tRNA genes reside on the same strand of the mtDNA. The gene sizes, positions and arrangements are depicted in Fig. 1.

Introns and intron-encoded ORFs. Eighty-one introns were found within the mitochondrial genome of *E. resinifera* WIN(M)79 and 72 of them contain ORFs coding for homing endonucleases (HEs). Among the 81 introns, according to RNAweasel²⁹, 12 can be assigned to group-IA, 32 to group-IB, 17 to group-IC, 8 to group-ID, 9 to group-I(derived) introns (Supplementary Table 1) and three introns have features diagnostic for group II introns³⁰⁻³². Seventy-two group I introns in *E. resinifera* encode one or two iORFs. In total, 76 LAGLIDADG and 15 GIY-YIG type ORFs were identified within the *E. resinifera* group I introns. Group II introns can code for either homing endonuclease or reverse transcriptases³¹. In this study we observed three group II introns. Two of them (located in *rns* and *rnl* gene) encoded LAGLIDADG type ORFs. Another group II intron (located in *cox1* gene) appeared to have no ORF.

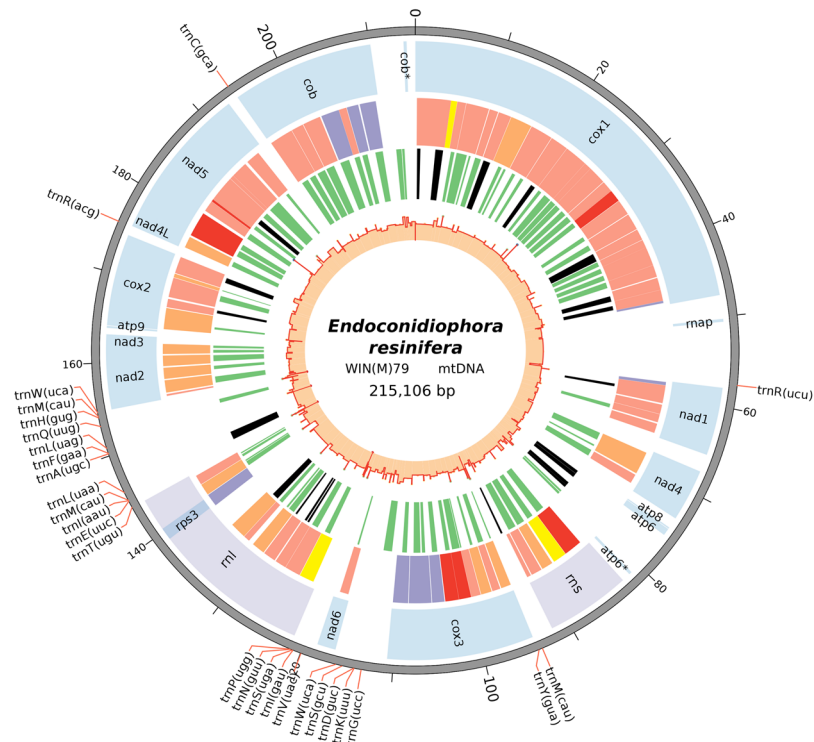


Figure 1. The annotated mitochondrial genome of *E. resinifera* [strain WIN(M)79]. The total size of this circular genome is 215 kb (represented by the scale). The position of the tRNAs are shown on the outer track, with the positions connecting to the scale with red lines. The first inner circle represents the position, size and the names of the protein-coding and rRNA genes. The introns are shown in the second inner circle and are colour coded according to the intron types/subtypes: group II (yellow), group IA (purple), group IB and group I derived (very light red), group IC (orange), and group ID (dark red). The third inner circle is to visualize the presence of the LAGLIDADG (green) or GIY-YIG (black) homing endonuclease genes encoded by the introns. The innermost circle is the GC plot of this genome; calculating GC% of genome features. Note the third inner circle also shows the location of free-standing homing endonuclease genes.

A tandem intron (mS917a and b) in the *rns* gene. Among the four strains of *E. resinifera*, the WIN1410B strain showed a unique type of intron arrangement within the *rns* gene (Fig. 2A). All strains analyzed for this species have a group ID intron at position S917 (intron insertion sites designated according to Johansen and Haugen³²). The other available *Ceratocystis* species also contain the mS917 intron. However, *E. resinifera* WIN(M)1410B has two group ID introns (instead of one) side by side (a and b) without any apparent exon sequence separating them. This special arrangement could be termed “tandem intron” or a “side by side” twintron; for consistency we will refer to this arrangement as a tandem intron. The mS917a and mS917b intron-encoded ORFs of the tandem intron are both related to the mS917 clade of LAGLIDADG ORFs previously characterized^{10,33}. Phylogenetic analysis showed that the components of the tandem intron mS917a and b ORFs are paralogues and other members of this clade can be located in group ID introns located within the *rnl*, *nad5*, *nad6* and *cox3* genes. Moreover, the intron ORF from the mS917b component groups with the intron ORFs located in the *rnl* intron (mL2029) clade, whereas the mS917a ORF groups with orthologues located within the *rns* mS917 intron. Based on the phylogenetic distribution of members of the 917 family of HEs it would appear that the mS917b intron/ORF is derived from a version of the mL2029 intron that has inserted (ectopically) immediately after the mS917a intron (Supplementary Fig. 1).

GC percentage and composition of the genome. The mitochondrial genome of *E. resinifera* is AT-rich (71%). The average GC content of the genome is 29% which is maintained across the mtDNA. The tRNA genes, however, have a higher GC (up to 50%) content compared to the rest of the genome. The genome is composed of genes (exons and introns), rRNAs (exons and introns), tRNAs, intron-encoded ORFs, and intergenic regions. However, the majority of the mitochondrial genome is comprised of introns and intron-encoded ORFs (68% of the genome). The introns embedded within the protein-coding genes make up 56% of the entire genome. The second major component of these genomes is comprised of the introns present within the rRNA genes (12%). The nucleotide sequences for protein-coding sequences (CDS) and rRNAs comprise 7% and 3% of the mitochondrial genome, respectively. The tRNAs make up 1% of the genome, and 21% of the mitochondrial genome is comprised of intergenic sequences. With regards to the 68% intron component, it can be arranged as follows for the intron subtypes: group IA, IB, IC, ID, I (derived) and group II as follows 14%, 42%, 22%, 9%, 10% and 3% respectively. The iORFs occupy half (50%) of the intron bases. The LAGLIDADG ORFs comprise 42% and the GIY-YIG ORFs make up 8% of all the bases that make up the introns (Supplementary Fig. 2).

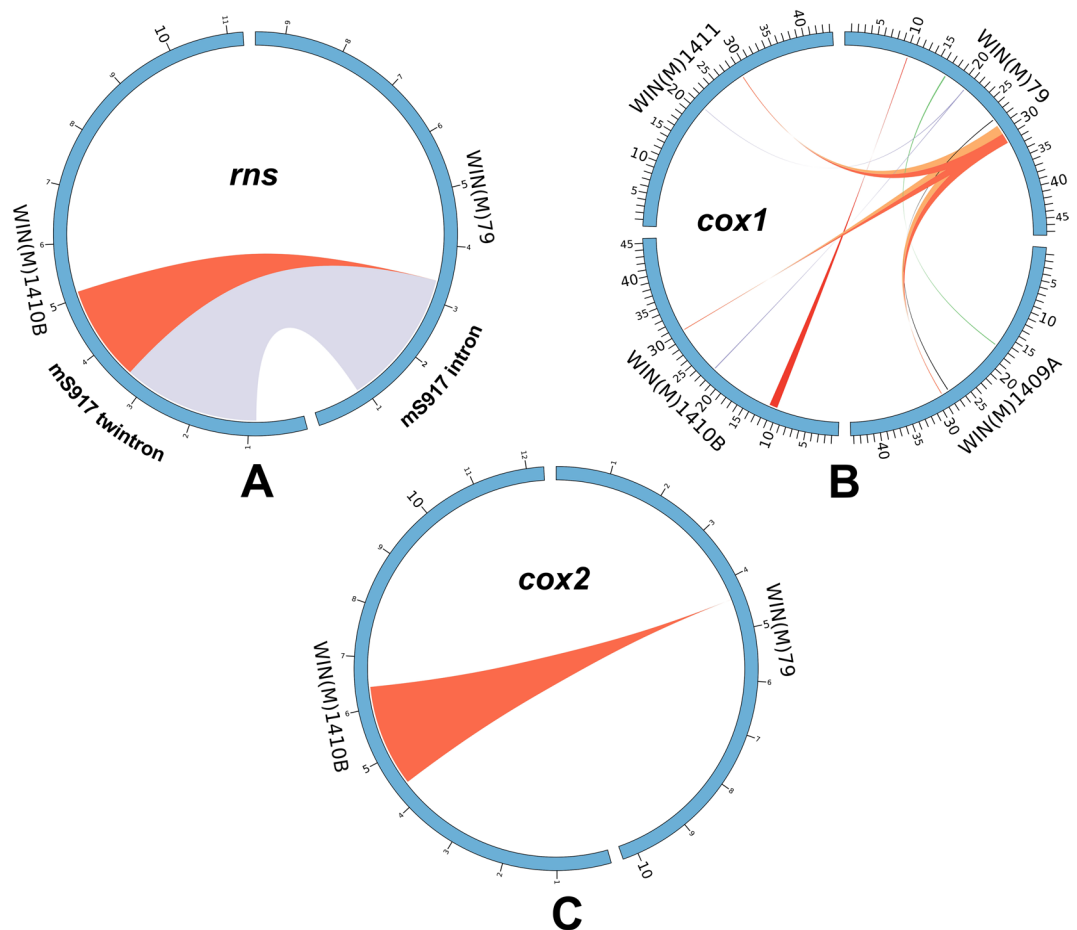


Figure 2. The comparison of *rns*, *cox1* and *cox2* genes from four strains [WIN(M)79, WIN(M)1409 A, WIN(M)1410B, WIN(M)1411] of *E. resinifera* considering the strain WIN(M)79 as a reference. (A) Comparing the *rns* gene showed that there is one novel group ID intron in WIN(M)1410B which is referred as a tandem intron (ms917). (B) Comparison of *cox1* genes showing that there are two additional introns in WIN(M)79 and one additional intron in WIN(M)1410B, moreover there are indels in different intronic regions. (C) With regards to the *cox2* gene WIN(M)1410B has one additional intron compared to the other strains.

The largest *cox1* gene recorded so far among the Ascomycota. The mitochondrial genome of *E. resinifera* is one of the largest [215 kb for WIN(M)79 and 220 kb for WIN(M)1410B] mitochondrial genomes so far reported for a member of the Ascomycota and it also contains the largest *cox1* gene (47.5 kb) recorded so far for any fungus. The size of this genome is for the most part due to the large number of introns (81 introns for WIN(M)79) and the *E. resinifera* mitochondrial genome in comparison to the other *Ceratocystis* spp. appears to have higher numbers of introns (Table 1). The *cox1* gene also appears to have expanded in *E. resinifera* to 47.5 kb due to the large number of introns (23 introns). Among the 23 introns, 22 are group I introns and one group II intron was identified. The group IB is the most abundant (17 group IB introns) intron type in the *cox1* gene (Supplementary Table 1). This gene appears to be rich in introns compared to other genes in the mitochondrial genome. The *cox1* gene is a very conserved gene (at the CDS level), but the intron numbers are variable among different strains examined for *E. resinifera* and also variable among species of *Ceratocystis* (Table 1). Among strains of *E. resinifera* *cox1* intron numbers are 21 for strains WIN(M)1409 A and 1411, 22 for strain WIN(M)1410B and 23 for strain WIN(M)79. This is in contrast to *C. platani*, *C. cacaofunesta* and *C. fimbriata* where the number of *cox1* introns is 9, 10 and 12, respectively.

Open Reading Frames and gene fragments within the intergenic spacers: HEGs and a plasmid-derived RNA polymerase. Twenty ORFs were detected in the intergenic regions of the mitochondrial genome of *E. resinifera*. A blastp search of those ORFs against the NCBI non-redundant database showed the presence of a partial DNA-dependent RNA polymerase (rnap) gene which showed similarity with a mitochondrial plasmid encoded rnap gene in *Neurospora intermedia*^{34,35}. Previously, a degenerated RNA pol gene was also reported from *C. cacaofunesta*. Nine degenerated GIY-YIG and eight LAGLIDADG (degenerated) ORFs along with partial duplications of the *atp6* and *cob* genes were also recorded from the intergenic spacers (Supplementary Table 2; Fig. 1). The partial duplication (C-terminal segment) of the *cob* gene is located downstream of the *cob* gene (genomic position: 213807–214152). There are three degenerated LAGLIDADG ORFs situated in the intergenic space between *cob* and the partial C-terminal duplication of the *cob* gene. The *atp6* gene is followed by a

| Organism | mtDNA size (bps) | GC% | Number of introns (per gene) | | | | | | | | | | | | | Total introns |
|--|------------------|-----|------------------------------|-------------|-------------|-------------|------------|-------------|-------------|------------|-------------|-------------|--------------|-------------|------------|-----------------|
| | | | <i>cox1</i> | <i>nad1</i> | <i>nad4</i> | <i>atp6</i> | <i>rns</i> | <i>cox3</i> | <i>nad6</i> | <i>rnl</i> | <i>nad2</i> | <i>cox2</i> | <i>nad4L</i> | <i>nad5</i> | <i>cob</i> | |
| <i>C. cacaofunesta</i> | 103,147 | 26 | 10 | 1 | 1 | 2 | 1 | 2 | 1 | 3 | 0 | 6 | 1 | 7 | 2 | 37 |
| <i>C. platani</i> | 116,162 | 27 | 9 | 2 | 1 | 2 | 2 | 4 | 0 | 5 | 1 | 5 | 1 | 6 | 3 | 41 |
| <i>C. fimbriata</i> | 141,204 | 27 | 12 | 3 | 1 | 2 | 1 | 4 | 0 | 4 | 2 | 6 | 1 | 10 | 5 | 51 |
| <i>E. resinifera</i> WIN(M)79 ^a (=UAMH 9644 ^b) | 215,106 | 29 | 23 | 5 | 2 | 0 | 5 | 9 | 1 | 11 | 5 | 5 | 1 | 7 | 7 | 81 |
| WIN(M)1409 A | 215,081 | 29 | 21 | 5 | 2 | 0 | 5 | 9 | 1 | 11 | 5 | 5 | 1 | 7 | 7 | 79 |
| WIN(M)1410B | 220,224 | 29 | 22 | 5 | 2 | 0 | 6c | 9 | 1 | 11 | 5 | 5 | 1 | 7 | 7 | 81 ^c |
| WIN(M)1411 | 214,750 | 29 | 21 | 5 | 2 | 0 | 5 | 9 | 1 | 11 | 5 | 5 | 1 | 7 | 7 | 79 |

Table 1. Comparison of the mitochondrial genomes of the *E. resinifera*, *C. cacaofunesta* (JX185564.1), *C. platani* (LBBL00000000.1) and *C. fimbriata* (APWK03000239.1). ^aWIN(M) culture collection of J. Reid, Department of Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada. ^bUAMH Centre for Global Microfungal Biodiversity, Division of Occupational & Environmental Health, Dalla Lana School of Public Health, University of Toronto, Toronto Ontario, Canada. ^cFor WIM(M)1410B the tandem intron located in the *rns* gene (mS917) was counted as two introns.

C-terminal duplication of the *atp6* gene and this duplicated segment is located between the *atp6* and *rns* genes (genomic position: 81091–81417). The duplicated *atp6* segment is flanked by 5 GIY-YIG and 2 LAGLIDADG type ORFs. Partial C-terminal duplications of the *cob* and *atp6* genes were noted in all four strains of *E. resinifera*. In strain WIN(M)1411 the complete duplication of the *trnA* gene was recorded and the duplicated version also included some of the upstream bases associated with the original copy of *trnA* (genomic position 149128–149198 duplicated at 149958–150028).

Degenerated *atp9*. The ATP synthase subunit 9 (*atp9*) gene sequence is present in the mitogenome of *E. resinifera* [genomic position: 163871–164096 in WIN(M)79] but it appears to have degenerated due to the presence of a premature stop codon. Blastx analysis showed a strong match (70% identity) to the *atp9* gene of *C. cacaofunesta* (GenBank accession YP_007507043.1). The same phenomenon was noted for *C. platani* (GenBank accession LBBL00000000.1) where its *atp9* gene sequence showed near 100% identity with the *atp9* sequence of *C. cacaofunesta*, but the *C. platani atp9* sequence was also interrupted by a premature stop codon. It is noteworthy that the *atp9* gene is absent in the *C. fimbriata* mitogenome (GenBank accession APWK03000239.1). Using the mitochondrial *atp9* amino acid sequence of *C. cacaofunesta* as a query against the *C. fimbriata* and *C. platani* genome data (including translated protein sequences) in a blastp analysis we noted that *atp9* sequences were located on the nuclear contig LBBL01000195.1 for *C. platani* and the nuclear scaffold APWK03000057.1 for *C. fimbriata*. The nuclear *atp9* gene products are also available from Genbank for *C. platani* (KKF93962.1) and *C. fimbriata* (PHH52759.1).

Mitochondrial genome comparison. The genomes of *E. resinifera* strains are highly conserved with polymorphism mainly due to the presence or absence of introns along with some short insertions and nucleotide substitutions in the non-coding sequences. The differences among the strains in gene sequences (including intron and exon) are compiled in Supplementary Table 3. Briefly, it was found that the CDS of gene sequences were highly conserved among the strains; only one silent mutation was noted in the *cox1* gene in the WIN(M)1410B strain (Supplementary Fig. 3). The *rnl* gene of WIN(M)1409 A and WIN(M)1410B showed small insertions and the *trnA* gene of WIN(M)1411 showed a small insertion (see Supplementary Table 3). But sources of mitochondrial genome variability among the strains are due to variations of the number of introns along with some small indels in the intronic sequences. Additional introns were noted in the *rns*, *cox1* and *cox2* genes of strain WIN(M)1410B (Fig. 2A,B). Moreover, two additional introns were found in the *cox1* gene of strain WIN(M)79 (Fig. 2B). We found no significant variations with regards to intergenic regions.

For a more detailed comparison among *Ceratocystis sensu lato* species, we have collected the mitochondrial genome sequences for *C. cacaofunesta*, *C. platani* and *C. fimbriata*. *Ceratocystis cacaofunesta* is fully annotated (GenBank accession: JX185564.1), and *C. platani* is available as one contig (but was not annotated) in GenBank (LBBL00000000.1) and in ENA database (GCA_000978885.1). The mitochondrial genome for *C. fimbriata* is available in GenBank as a single contig (APWK03000239.1) but also not annotated. We have annotated *C. platani* and *C. fimbriata* for this study. The protein-coding regions were translated and extracted to compile a concatenated dataset that allowed these fungi to be compared with each other along with other members of the Ascomycota. The phylogenetic analysis showed that the *Ceratocystis s.l.* spp. are distinct from each other and they do comprise a separate clade (Microscales) in the phylogenetic tree based on concatenated mtDNA protein-coding sequences (Supplementary Fig. 1). All four members of *Ceratocystis s.l.* grouped into one clade with *E. resinifera* forming the basal member and sequences for *C. platani* and *C. fimbriata* grouping together. The phylogenetic tree overall showed strong node support values for the major nodes and the Microscales were positioned between the following Orders Hypocreales and Glomerellales (Supplementary Fig. 4).

The variation in genome sizes, GC content and the presence of introns in every gene among the *Ceratocystis s.l.* species are listed in Table 1. The data showed that the genome size and intron number of *E. resinifera* is about

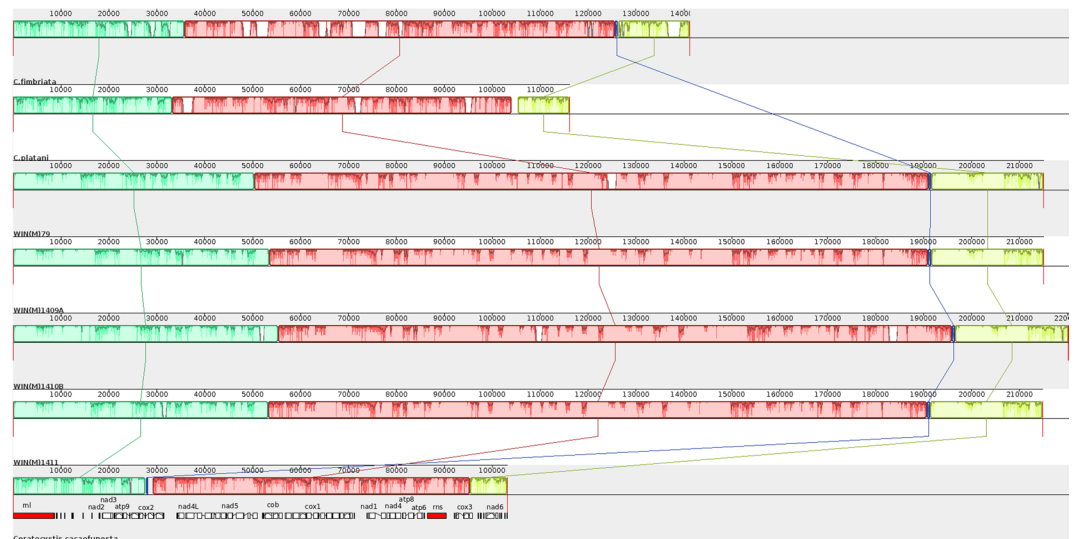


Figure 3. Genome-wide comparison for species of *Ceratocystis* Mauve. The progressiveMauve alignment (in Mauve program) shows the homologous blocks shared among the mitochondrial genomes and it also connected these blocks with lines, indicating corresponding position among the homologous blocks in order to visualize the gene arrangement.

double compared to other *Ceratocystis* species. A comparative alignment of all these mitochondrial genomes was done with the Mauve program (Fig. 3). It clearly showed the homologous blocks shared between these genomes and it also showed a linear relationship among the genes, which implies that the gene order or synteny of these genomes is conserved. Besides the gene synteny for protein and rRNA coding genes, the comparison of the tRNA genes also showed conservation of gene order. However, the number of tRNA genes is not the same among the examined species: 25 tRNAs in *C. fimbriata*, 26 in *C. platani*, 27 in *E. resinifera* and 30 in *C. cacaofunesta* (Supplementary Table 4).

Discussion

Mitochondrial genome architecture among members of *Ceratocystis sensu lato*. The phylogenetic tree generated for ascomycetes fungi, based on concatenated mitochondrial protein sequences, generated a well-supported topology consistent with the topologies of previously published reports based on rDNA data^{2,36}. *Ceratocystis* and allied taxa belong to the Microascales and are distinct from species that can be assigned to other orders such as the Hypocreales, Glomerellales, Xylariales, Sordariales, Ophiostomatales etc. The assembly of the mitogenomes for the tested strains of *E. resinifera* can be represented as circular molecules ranging in size from 214,750 to 220,224 nucleotides. It should be noted that fungal mtDNAs could also have linear topologies and have been proposed to occur as long concatemers, possibly products of a recombination-dependent rolling circle-type DNA replication mechanism^{37–39}.

The mitochondrial genomes of *E. resinifera* are the largest genomes reported so far for members of the Ascomycota; yet with regards to the standard mtDNA core genes these genomes do not offer additional genes compared to other fungal mitochondrial genomes^{40–42}. The *E. resinifera* mitochondrial genome contains 15 protein coding genes, 2 rRNA genes and 27 tRNA genes similar to other fungal mitochondrial genomes. Moreover, the *rps3* gene is embedded within an *rnl* group IA type intron (mL2449), which is a common feature in many filamentous ascomycetes fungi mitogenomes⁴³. The *atp9* gene apparently is found to be present in some fungi and but not in others^{44,45}. With regards to the mitochondrial genomes examined in this study the *atp9* gene is present in *C. cacaofunesta*, but in *E. resinifera* and *C. platani* the *atp9* gene has accumulated mutations that generated a premature stop codon and in *C. fimbriata* the *atp9* gene is missing. This would suggest that the *atp9* gene is drifting in some species and a nuclear counterpart might be available that can compensate for the loss of the mitochondrial version of the *atp9* gene. Indeed examining the nuclear genomes of *C. platani* and *C. fimbriata* uncovered nuclear encoded versions of *atp9* suggesting that a copy of the mtDNA *atp9* gene has been transferred to the nuclear genome in these fungi. Similar findings were noted for *Stemphylium lycopersici* where *atp9* is missing from the mitochondrial genome but a complete version could be located in the nuclear genome⁴⁴.

The progressiveMauve alignment of the mitochondrial genomes for the *Ceratocystis s.l.* species showed that gene synteny is conserved and variations in mtDNA and gene sizes are mostly due to the expanding numbers of introns. Variation among various strains of *E. resinifera* is restricted to one synonymous substitution in the *cox1* gene, a few SNPs within the intronic sequences or other non-CDS bases along with a few indels within the intergenic regions. Similar to what has been observed in other fungi such as *Chrysosporthe* species⁴⁶, *Aspergillus* and *Penicillium* species²¹, and *Saccharomyces sensu stricto* species⁴⁷, we observed intron derived size polymorphism among strains of *E. resinifera*.

Other noteworthy features are the fusions of several gene pairs typically involving a one nucleotide overlap among the two reading frames. The overlap of *nad2* with *nad3* genes and *nad4L* with *nad5* by one nucleotide has been noted in other fungi²².

Mobile elements and genome expansion (duplication and degeneration). Recent papers have noted that fungal mitochondrial genomes are dynamic with regards to their structure and composition due to the presence of mobile elements (such as group I and group II introns) and duplication events^{20,42,48}. This study found 81 introns in the mtDNAs of *E. resinifera* strains examined which is nearly double the number of introns compared to the other species of *Ceratocystis*. Most noteworthy is the *cox1* gene from *E. resinifera* strain WIN(M)79 that has 23 introns and this gene is 45.7 kb long. The size of the *cox1* gene alone exceeds the sizes of many complete fungal mitochondrial genomes⁴⁹. The *E. resinifera cox1* gene is the largest reported so far, previously the *cox1* gene from *Agaricus bisporus* at ~30 kb long with 19 introns was reported to be the largest *cox1* gene among the fungi⁵⁰. In *E. resinifera* the *cox1* gene is considerable longer and acquired 4 more introns; combined with the intron encoded ORFs this expanded the size of the gene to 45.7 kb. The *cox1* gene has been utilized as a DNA barcoding marker in metazoans, but the presence of potentially large numbers of introns makes the *cox1* gene not very suitable for fungal DNA- based barcoding⁵¹. Mobile elements that require specific target sequences such as group I and group II introns favor genes that are under functional constraints and are highly conserved, making intra- and intergenomic mobility more feasible.

Examples of degenerated intron ORFs were noted and these are to be expected as according to Goddard and Burt²⁷ introns and encoded ORFs such as homing endonucleases are not subject to natural selection, thus their sequences drift and can accumulate deleterious mutations. Neutral evolution is thus a plausible model to explain the potential genome expansion noted among some members of *Ceratocystis s.l.* Although introns appear to be the major factor that contributes towards mtDNA size expansion in *E. resinifera*, insertion of plasmid components (such as the *rns* gene), and gene duplication events (partial duplication of *atp6*, *cob* and HEGs) and the expansion of intergenic spacers also contribute towards the size of the mitochondrial genome.

Overall, the examined mitochondrial genomes for *Ceratocystis* and *Endoconidiophora* species appear to evolve rapidly in gene structure (i.e. intron composition) but slowly in sequence and gene order. This has also been observed in plant mitochondrial genomes and some fungi^{46,52}. Therefore, our findings show that mtDNA polymorphisms are mostly due to the presence and absence of introns.

Tandem intron located at mS917. So called twintrons have been described from various fungal mitochondrial genomes with various combinations of group I or group II introns nested inside each other. These elements may require that during RNA processing the internal member has to splice first before the external member can be excised from the transcript⁵³. Deng *et al.*⁴⁹ noted that in *Hypomyces aurantius* the *cox3* gene harbored a twintron (*cox3-i2*) that is a “side-by-side twintron” where two group IA introns are arranged in tandem. The *rns* gene in *E. resinifera* [WIN(M)1410B] contains a twintron where two group ID introns are placed next to each other at the S917 position of the *rns* gene. This position (S917) has previously been noted to be invaded in some fungi by a group ID intron that expresses active HEases³⁴, in addition it has been recorded that this location in *Cryphonectria parasitica* can be occupied by a twintron where a group ID intron that encodes a double motif LAGLIDADG-type ORF is inserted into an ORF-less external group ID intron^{11,54}. This arrangement differs from that observed in *E. resinifera* (strain WIN(M)1410B) where two LAGLIDADG ORF-encoding group ID introns are situated next to each other. Based on the phylogenetic relationships between the two members of this tandem intron, it appears the 5' member is the original resident of the S917 site and the 3' component is due to an ectopic integration event whereby a paralog of the mS917 HEG which was probably located in the *rnl* gene reinvaded the mS917 position. This intron arrangement warrants further characterization in future studies with regards to its splicing pattern and the target preferences for the intron encoded HEases.

Evolutionary dynamics of the introns and HEGs and the mitochondrial genome. Introns comprise 68% of the mitochondrial genome in *E. resinifera*, and most of the introns contain ORFs encoding putative homing endonucleases. Those ORFs comprise 50% of the size of the introns. Group I introns can move to cognate alleles that lack introns or in some instances, ectopically integrate into new sites, as they encode homing endonucleases²⁵. Intron-loss can be mediated when a reverse-transcribed mature transcript replaces the original intron-containing gene⁵⁵. Deletion of introns can also be due to intra- or intergenomic recombination events⁵⁶. The evolutionary dynamics of introns and homing endonucleases is quite complex, the gain and loss of introns and their encoded ORFs tends to be attributed to a HEG lifecycle²⁷ that is based on neutral evolution. The model is based on the observation that among members of the Saccharomycetales the omega intron (*rnl* gene introns) appears to undergo a cycle of invasion and degeneration; as there is no selection, introns and encoded ORFs accumulate mutations that eventually lead to their erosion and loss. To persist these elements have to keep invading cognate intron-minus alleles, transpose into new sites or move horizontally into new genomes, or gain a beneficial function^{27,28}.

The large number of introns noted in some fungal mitochondrial genomes such as *Agaricus bisporus*, *Rhizoctonia solani*, *Cryphonectria parasitica*, *Sclerotinia borealis*, *E. resinifera* etc. are in contrast to small fungal mtDNAs encoding only one intron such as *Sporothrix schenckii* and *S. globosa*; this raises the question if drift is indeed the only possible explanation for the distribution of introns within fungal mitochondrial genomes? Also, some introns appear to be quite conserved such as the *rnl* intron that in many ascomycetes fungi encodes *rps3*. It is assumed that encoding a potentially essential gene ensures these introns are not subject to neutral evolution. Other introns for less obvious reason also appear to be observed at relative higher frequencies compared to other introns, such as *cob393* and *cob490* and possibly these introns may provide some means for gene regulation and therefore selection may favour their maintenance⁵⁷. Rudan *et al.*⁵⁸ recently presented data from *S. cerevisiae* that

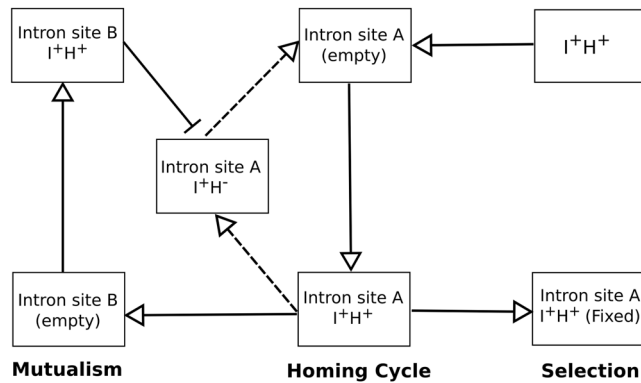


Figure 4. The fate of composite elements such as introns plus IEPs ($I+H+$). The composite element invades an empty site and from here it could spread into other sites (only site B shown for simplicity) and related IEPs could still interact with their ancestral intron version possibly facilitating splicing or mobility thus enhancing the chances of the ancestral intron to persist. This form of mutualism could even complement situations where the ancestral intron ORF has started to accumulate deleterious mutations ($H-$). Other composite elements may be strictly subject to drift and first the ORF is degenerating and eventually the intron is lost from the genome and possibly from the population. There might be situations where the composite elements have been co-opted as maturases or regulatory elements or as platforms for expressing essential genes (*rps3*) and these introns would be subject to adaptive selection and thus could become fixed in the genome and the population (Figure adapted from Gogarten and Hilario⁸³; licensed under a CCBY 2.0 license, (<https://creativecommons.org/licenses/by/2.0/>)). This Figure adds a new component to the standard homing endonucleases life cycle^{27,83} that suggests that some HEGs can avoid extinction due to mutualistic interactions that can complement for the accumulation of mutations.

suggests the mtDNA introns are important in fine-tuning gene expression and they facilitate the generation of appropriate amounts of transcripts. Belfort⁵⁹ has suggested that some self-splicing elements could be bio-sensors that can modulate the expression of the genes that contain introns or inteins. Conversely, many introns (intron insertion sites) have a rather disjointed distribution among the fungi encoding ORFs at various state of degeneration; these introns may be excellent examples of neutral evolution as proposed by Goddard and Burt²⁷.

Another category of introns is represented by those that appear to have invaded new sites within the same genome⁶⁰, a temporary means of escaping the Goddard and Burt²⁷ HEG lifecycle of invasion, decay and eventual loss. Related introns present within the same genome may still interact *in trans* in some collaborative fashion rendering them less prone to extinction^{61,62}. Some intron encoded proteins (IEPs) can also act as maturases that facilitate the intron RNA to fold into a splicing-competent configuration²⁹. In *S. cerevisiae*, two homologous IEPs have been characterized, and *cyt b14* is required for splicing of both the *cyt b14* intron and the *cox1 a14* intron but the *cox1 a14* IEP can generate double stranded cuts within *cox1* sequences⁶³. Trans-acting interactions between introns and free standing HEGs have been noted among phages; collaborative homing refers to scenarios where a HE can catalyze the mobility of an intron as both share the same insertion/target site^{64–66}. In scenarios such as the mS917 clade of introns where orthologous elements have spread into the *rns*, *rnl*, *nad5*, *nad6* and *cox3* genes³⁴ one can propose that there might be some interactions among members of this clade. *E. resinifera* has mS917 members located within the *rns*, *rnl*, *nad5* and *nad6* genes in addition one strain (WIN(M)1410B) contains a tandem intron located at mS917 composed of two members (*rns* and *rnl* version, Fig. 2A) of this family of introns. When members of a HEG family are present within the same genome one can envision a “hypercycle”-like analogy^{67,68}; however, here dependencies are more loose as individual members can drift, become selfish and some members can be short circuited⁶⁹. Members would not have to interact in a directional manner instead interactions such as trans-acting maturase activities or trans-acting homing activity mostly likely would be linked to those that recently diverged from each other. This arrangement would provide some degree of stability for the persistence of members of a HEG family allowing members to maintain their numbers or even spread to new locations “outpacing” drift as predicted by the Goddard and Burt model²⁷. These types of interactions (Fig. 4) may in part explain why some mitochondrial genomes have expanded by gaining or maintaining large numbers of introns.

Horizontal gene transfer, inter and intra-genome intron mobility, plus gene conversion promoted by IEPs and host genome repair systems combined with drift are the mechanisms that appear to promote intron diversity and potentially intron expansion in fungi^{20,60,69,70}. Why are some fungal genomes almost completely devoid of introns? This could be again due to drift and the biased transmission of mtDNAs that are reduced in size or replicative advantage of smaller mtDNAs or loss of mtDNA introns could be the result of nuclear genome rearrangements that result in the loss of factors that can promote intron splicing, intron RNA stability, intron homing or mtDNA DNA repair (recombination).

Conclusion

So far, few mitochondrial genomes are currently available for members of the Microascales. This study examines the mitochondrial genome of *E. resinifera* a species that used to be assigned to the genus *Ceratocystis*. The latter is a genus that has recently undergone extensive taxonomic revisions⁷¹ and the mitochondrial genome might offer mtDNA loci that could be developed into molecular markers assisting in the identification of taxa within

this group of economically important fungi. Large mitochondrial genomes offer some insight on mechanisms that might cause these genomes to expand. With regards to *E. resinifera*, introns appear to be a major contributor towards genome expansion. Introns and their encoded homing endonucleases all assemble and initiate further invasions by drift⁵⁷ but once they have inserted into a gene several mechanisms may determine their fate. Many probably do fit the model proposed by Goddard and Burt²⁷ whereby these elements drift and thus face eventual elimination due to the accumulation of mutations and persistence within a population requires re-invasion of new loci or loci that lost introns. However, there appears to be evidence that some introns may actually be beneficial either encoding proteins that are useful to the genome (maturase activity, *rps3*) or introns that can act as gene regulators and thus these introns are maintained within a population. Finally, some introns appear to persist as they are co-operating with other introns promoting a system of mutualism that renders them less prone to extinction.

Material and Methods

Culturing fungi. The fungi were grown at 25 °C for 8–10 days on malt extract agar (MEA - 3% malt extract, 2% agar and 0.1% yeast extract) plates. Mycelium was scraped from these plates and transferred to 1 L flasks containing 250 mL yeast extract, peptone, dextrose broth (YPD - 0.1% yeast extract, 0.1% peptone, 0.3% dextrose). The YPD broth cultures were maintained for 8–10 days at 25 °C. The fungal strains of *E. resinifera* utilized in the study are listed in Table 1.

Isolation of Mitochondria. Fungal mycelia was collected by vacuum filtration using a Büchner funnel and Whatman[®] qualitative filter paper. The mycelium was disrupted by grinding with mortar and pestle with the addition of 2 mL of isolation buffer [10 mM Tris-HCl (pH 8.0), 440 mM sucrose, 5 mM ethylene-diamine-tetra-acetic acid (EDTA)] and 1.5 g of acid-washed sand for each 1 g of mycelia. The fungal material was ground for about 5–10 minutes until the mycelia/sand/buffer mixtures forms a slurry. This slurry was transferred to a 25 mL Corex[®] centrifuge tube (ThermoFisher) and centrifuged for 15 min at 3000 g using a Sorvall[®] SS-34 fixed angle rotor in a Sorvall[®] RC-5B Plus centrifuge at 4 °C to pellet nuclei, cell debris and sand. The supernatant was transferred to a 25 mL Corex[®] centrifuge tube (ThermoFisher) and centrifuged at 20 000 g using a Sorvall[®] SS-34 fixed angle rotor in a Sorvall[®] RC-5B Plus centrifuge at 4 °C for 30 min to pellet the mitochondria.

Mitochondrial DNA extraction. The mitochondrial enriched pellet was suspended in 3 mL of extraction buffer [100 mM Tris-HCl (pH 8.0), 2% cetyl-trimethylammonium bromide (CTAB), 20 mM EDTA, 1.4 M NaCl] plus 330 µL of 20% sodium dodecyl sulfate (SDS) and nucleic acids were extracted based on Hausner *et al.*⁷². Briefly the mixture was incubated for 2 hours (or overnight) at 55–65 °C and proteins and lipids were removed by adding an equal volume (~3 mL) of chloroform, after mixing the contents of the tube it was centrifuged at 3000 g with a IEC Centra CL2 centrifuge for 5 min. The top aqueous layer was transferred to a new 15 mL centrifuge tube and mixed with 4 µL of RNase A (QIAGEN) and incubated in a 55–65 °C water bath for 1 h to remove RNA. RNase was removed by addition of Chloroform in a 1:1 ratio in the tube and centrifuged at 3000 g for 20 min in an IEC Centra CL2 centrifuge. The aqueous layer was transferred to a new tube and mixed with 2.5 volumes of 95% ethanol and placed in the freezer at -20 °C for 1 hour. The mixture was then centrifuged at 3000 g for 15 min in an IEC Centra CL2 centrifuge to pellet the DNA. Supernatant was removed and the DNA pellet was washed with 1 mL of 70% ethanol, the tube was centrifuged again at 3000 g for 5 min and the ethanol was removed. The DNA pellet was air dried and suspended in 200 µL DNase/RNase-free water and placed in -20 °C for storage.

Quantifying DNA. The extracted DNA was quantified with a NanoDrop 2000c UV-Vis Spectrophotometer and the quality was determined on the basis of the 260/280 and 260/230 OD ratio. Quantification was confirmed by gel electrophoresis of 10 µL of the extracted DNA sample on a 1% agarose gel.

Genome sequencing and assembly. The mitochondrial genomic DNA was sequenced and assembled by Génome Québec (Innovation Centre, McGill University). For each sample 75 µL of DNA (~1 µg) supplied within an Eppendorf 96-well twin.tec[®] PCR plate (Cat. No. 951020401) sealed with VWR[®] aluminum foil (Cat. No. 60941-074), was sent to Génome Québec for Illumina sequencing using the MiSeq platform. The DNAs from different fungal samples were barcoded and combined into a single MiSeq run. The quality of the sequence reads were verified by FastQC⁷³. The sequenced reads generated from NGS sequencing were assembled by the a5-miseq-pipeline⁷⁴ – a MiSeq optimization of the original a5 pipeline⁷⁵.

Genome Annotation. The assemblage of the genome by a5-miseq pipeline yielded a set of scaffolds. The scaffolds were sorted out on the basis of the scaffold-size and presence of mitochondrial genes and those scaffolds were used to join and construct an entire uninterrupted mitochondrial genome sequence by using custom python script (available upon request) and NCBI-blast + program as well as by manual inspection. The position of the protein-coding genes, rRNAs and tRNAs were identified by MFannot⁷⁶. tRNA genes were identified with tRNAscan-SE⁷⁷. Intron-exon junctions within protein-coding and rRNA genes were initially obtained in MFannot and verified by multiple sequence alignments (MSA) of a gene and aligning it to the CDS of the same gene from related species. Sequence alignments were performed with MAFFT⁷⁸. Sequences were also analyzed with the RNAweasel program²⁹ to determine intron types and subtypes. Intron sequences were also examined with the ORF-Finder program (NCBI) to identify possible ORFs. Further, the Smart-BLAST program was used to determine the type of the intron-encoded ORFs. The coordinates of all genes, rRNA, tRNA, introns and intron-encoded ORFs and any other features were annotated using Artemis⁷⁹ and visualized in Circos⁸⁰.

Genome comparison. The annotated *E. resinifera* mitochondrial genomes were compared for their features by generating multiple sequence alignments (MSA). The MSA allowed for noting SNPs, indels and

polymorphisms that relate to the presence and absence of introns. For comparative purposes the mitochondrial genomes of *C. cacaofunesta*, *C. platani* and *C. fimbriata* were also included in the MSA.

A phylogenetic tree, based on a concatenated data set of 13 protein sequences (in alphabetical order: *atp6*, *atp8*, *cob*, *cox1-3*, *nad1-6*) was constructed in MrBayes⁸¹ based on an alignment of 41 fungal species generated with MAFFT. The tree topology generated by MrBayes was compared with previously published phylogenetic trees based on rDNA data^{2,36} and they were found to be consistent with one another. Also, a comparative map was generated in Mauve⁸² to visualize variations in the genomic architecture.

Data Availability Statement

All data nucleotide sequences generated are available in GenBank also data generated and analysed during this study are included in this published article (and its Supplementary Information files).

References

- Behrendt, C. J., Blanchette, R. A. & Farrell, R. L. Biological control of blue-stain fungi in wood. *Phytopathology* **85**, 92–97 (1995).
- de Beer, Z. W., Duong, T. A., Barnes, I., Wingfield, B. D. & Wingfield, M. J. Redefining *Ceratocystis* and allied genera. *Stud Mycol.* **79**, 187–219 (2014).
- Halsted, B. D. & Fairchild, D. G. Sweet-potato black rot. *J Mycol.* **7**, 1–11 (1891).
- Juzwik, J., Harrington, T. C. & MacDonald, W. L. The origin of *Ceratocystis fagacearum*, the oak wilt fungus. *Annual Review of Phytopathology.* **46**, 13–26 (2008).
- Engelbrecht, C. J., Harrington, T. C. & Alfenas, A. *Ceratocystis* wilt of cacao — a disease of increasing importance. *Phytopathology.* **97**, 1648–1649 (2007).
- Tsoelas, P., Santini, A., Wingfield, M. J. & de Beer, Z. W. Canker stain: A lethal disease destroying iconic plane trees. *Plant Disease* **101**(5), 645–658, <https://doi.org/10.1094/PDIS-09-16-1235-FE> (2017).
- Houston, D. R. Recognizing and managing sapstreak disease of sugar maple. *Res Pap. NE* **675**, 1–11 (1993).
- Harrington, T. C. & Wingfield, M. J. The *Ceratocystis* species on conifers. *Can J Bot.* **76**, 1446–1457 (1998).
- Wingfield, M. J., Seifert, K. A. & Webber, J. F. *Ceratocystis* and *Ophiostoma* – Taxonomy, ecology and pathology (eds Wingfield, M. J., Seifert, K. A. & Webber, J. F.) (The American Phytopathological Society Press, 1993).
- Ambrosio, A. B. *et al.* Global analyses of *Ceratocystis cacaofunesta* mitochondria: from genome to proteome. *BMC Genomics* **14**, 91, <https://doi.org/10.1186/1471-2164-14-91> (2013).
- Hafez, M. *et al.* The mtDNA rns gene landscape in the Ophiostomatales and other fungal taxa: twintrons, introns, and intron-encoded proteins. *Fungal Genet. Biol.* **53**, 71–83 (2013).
- Sethuraman, J. *et al.* Evolutionary dynamics of introns and their open reading frames in the U7 region of the mitochondrial rnl gene in species of *Ceratocystis*. *Fungal Biol.* **117**(11–12), 791–806 (2013).
- Wilken, P. M., Steenkamp, E. T., Wingfield, M. J., De Beer, Z. W. & Wingfield, B. D. IMA Genome-F 1: Draft nuclear genome sequence for the plant pathogen, *Ceratocystis fimbriata*. *IMA Fungus.* **4**, 357–358 (2013).
- Van der Nest, M. A. *et al.* Draft genome sequences of *Diplodia sapinea*, *Ceratocystis manginecans*, and *Ceratocystis moniliformis*. *IMA Fungus* **5**, 135–140 (2014).
- Wingfield, B. D. *et al.* IMA Genome-F5: Draft genome sequences of *Armillaria fuscipes*, *Ceratocystiopsis minuta*, *Ceratocystis adiposa*, *Endoconidiophora laricicola*, *E. polonica* and *Penicillium frei* DAOMC 242723. *IMA Fungus* **7**, 217–227 (2016).
- Wingfield, B. D. *et al.* IMA Genome-F 7: Draft genome sequences for *Ceratocystis fagacearum*, *C. harringtonii*, *Grossmannia penicillata*, and *Huntia bhutanensis*. *IMA Fungus* **7**, 317–323 (2016).
- Hausner, G. Fungal mitochondrial genomes, plasmids and introns in *Applied Mycology and Biotechnology Volume III: Fungal Genomics* (eds Arora, D. K. & Khachatourians, G. G.) 101–131 (Elsevier Science, 2003).
- Freel, K. C., Friedrich, A. & Schacherer, J. Mitochondrial genome evolution in yeasts: an all-encompassing view. *FEMS Yeast Res.* **15**(4), fov023, <https://doi.org/10.1093/femsyr/fov023> (2015).
- James, T. Y. *et al.* Shared signatures of parasitism and phylogenomics unite Cryptomycota and microsporidia. *Curr Biol.* **23**(16), 1548–53, <https://doi.org/10.1016/j.cub.2013.06.057> (2013).
- Losada, L. *et al.* Mobile elements and mitochondrial genome expansion in the soil fungus and potato pathogen *Rhizoctonia solani* AG-3. *FEMS Microbiol Lett.* **352**(2), 165–73 (2014).
- Joardar, V. *et al.* Sequencing of mitochondrial genomes of nine *Aspergillus* and *Penicillium* species identifies mobile introns and accessory genes as main sources of genome size variability. *BMC Genomics* **13**, 698, <https://doi.org/10.1186/1471-2164-13-698> (2012).
- Aguileta, G. *et al.* High variability of mitochondrial gene order among fungi. *Genome Biol Evol.* **6**(2), 451–465 (2014).
- Lang, B. F., Laforest, M. J. & Burger, G. Mitochondrial introns: a critical view. *Trends Genet.* **23**(3), 119–25 (2007).
- Belfort, M., Derbyshire, V., Parker, M. M., Cousineau, B. & Lambowitz, A. M. Mobile introns: pathways and proteins in *Mobile DNA II* (eds Craig, N. L., Craigie, R., Gellert, M., Lambowitz, A. M.) 761–783 (ASM Press, 2002).
- Belfort, M. Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. *Genes Dev.* **17**, 2860–2863 (2003).
- Edgell, D. R., Chalamcharla, V. R. & Belfort, M. Learning to live together: mutualism between self-splicing introns and their hosts. *BMC Biol.* **9** (2011).
- Goddard, M. R. & Burt, A. Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci USA* **96**(24), 13880–5 (1999).
- Wu, B. & Hao, W. Horizontal transfer and gene conversion as an important driving force in shaping the landscape of mitochondrial introns. *G3 (Bethesda)* **4**, 605–612 (2014).
- Beck, N. & Lang, B. F. RNAweasel, a webserver for identification of mitochondrial, structured RNAs. <http://megasun.bch.umontreal.ca/cgi-bin/RNAweasel/RNAweaselInterface.pl> (2009).
- Michel, F. & Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol.* **216**(3), 585–610 (1990).
- Toor, N. & Zimmerly, S. Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. *RNA* **8**, 1373–1377 (2002).
- Hausner, G., Hafez, M. & Edgell, D. R. Bacterial group I introns: mobile RNA catalysts. *Mobile DNA* **5**, 8, <https://doi.org/10.1186/1759-8753-5-8> (2014).
- Johansen, S. & Haugen, P. A new nomenclature of group I introns in ribosomal DNA. *RNA* **7**(7), 935–936 (2001).
- Bilto, I. M., Guha, T. K., Wai, A. & Hausner, G. Three new active members of the I-OnuI family of homing endonucleases. *Can J Microbiol.* **63**(8), 671–681 (2017).
- Xu, Y., Yang, S., Turitsa, I. & Griffiths, A. Divergence of a linear and a circular plasmid in disjunct natural isolates of the fungus *Neurospora*. *Plasmid* **42**(2), 115–25 (1999).
- Schoch, C. L. *et al.* The Ascomycota tree of life: a phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. *Syst Biol.* **58**(2), 224–239 (2009).

37. Bendich, A. J. Reaching for the ring: the study of mitochondrial genome structure. *Curr Genet.* **24**(4), 279–90 (1993).
38. Baidyaroy, D., Hausner, G. & Bertrand, H. *In vivo* conformation and replication intermediates of circular mitochondrial plasmids in *Neurospora* and *Cryphonectria parasitica*. *Fungal Biol.* **116**(8), 919–31 (2012).
39. Hausner, G., Nummy, K. A. & Bertrand, H. Asexual transmission, non-suppressiveness and meiotic extinction of small plasmid-like derivatives of the mitochondrial DNA in *Neurospora crassa*. *Fungal Genet Biol.* **43**(2), 90–101 (2006).
40. Salavirta, H. 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.* **9**(5), e97141, <https://doi.org/10.1371/journal.pone.0097141> (2014).
41. Kang, X., Hu, L., Shen, P., Li, R. & Liu, D. SMRT sequencing revealed mitogenome characteristics and mitogenome-wide DNA modification pattern in *Ophiocordyceps sinensis*. *Front Microbiol.* **8**, 1422, <https://doi.org/10.3389/fmicb.2017.01422> (2017).
42. Mardanov, A. V. *et al.* The 203 kbp mitochondrial genome of the phytopathogenic fungus *Sclerotinia borealis* reveals multiple invasions of introns and genomic duplications. *PLoS One.* **9**(9), e107536, <https://doi.org/10.1371/journal.pone.0107536> (2014).
43. Sethuraman, J., Majer, A., Iranpour, M. & Hausner, G. Molecular evolution of the mtDNA encoded rps3 gene among filamentous ascomycetes fungi with an emphasis on the Ophiostomatoid fungi. *J Mol Evol.* **69**(4), 372–85 (2009).
44. Franco, M. E. E. *et al.* The mitochondrial genome of the plant-pathogenic fungus *Stemphylium lycopersici* uncovers a dynamic structure due to repetitive and mobile elements. *PLoS One.* **12**(10), e0185545, <https://doi.org/10.1371/journal.pone.0185545> (2017).
45. van de Sande, W. W. Phylogenetic analysis of the complete mitochondrial genome of *Madurella mycetomatis* confirms its taxonomic position within the order Sordariales. *PLoS One.* **7**(6), e38654, <https://doi.org/10.1371/journal.pone.0038654> (2012).
46. Kanzi, A. M. *et al.* Intron derived size polymorphism in the mitochondrial genomes of closely related *Chrysosporthe* species. *PLoS One* **11**(6), e0156104, <https://doi.org/10.1371/journal.pone.0156104> (2016).
47. Ruan, J., Cheng, J., Zhang, T. & Jiang, H. Mitochondrial genome evolution in the *Saccharomyces sensu stricto* complex. *PLoS One* **12**(8), e0183035, <https://doi.org/10.1371/journal.pone.0183035>.
48. Jalalzadeh, B. *et al.* The interspecific variability of mitochondrial genes of *Agaricus bisporus* reveals an extensive group I intron mobility combined with low nucleotide substitution rates. *Curr Genet.* **61**, 87–102 (2015).
49. Deng, Y. *et al.* Analysis of the mitochondrial genome in *Hypomyces aurantius* reveals a novel twintron complex in fungi. *Int J Mol Sci.* **17**(7), 1049, <https://doi.org/10.3390/ijms17071049> (2016).
50. Férandon, C. *et al.* The *Agaricus bisporus* cox1 gene: the longest mitochondrial gene and the largest reservoir of mitochondrial group I introns. *PLoS One.* **5**(11), e14048, <https://doi.org/10.1371/journal.pone.0014048> (2010).
51. Schoch, C. L. *et al.* Nuclear ribosomal internal spacer (ITS) region as a universal DNA barcode marker for fungi. *Proc Natl Acad Sci USA* **109**(16), 6241–6 (2012).
52. Palmer, J. D. *et al.* Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proc Natl Acad Sci USA* **97**(13), 6960–6 (2000).
53. Hafez, M. & Hausner, G. Convergent evolution of twintron-like configurations: One is never enough. *RNA Biol.* **12**(12), 1275–1288 (2015).
54. Monteiro-Vitorello, C. B. *et al.* The *Cryphonectria parasitica* mitochondrial rns gene: plasmid-like elements, introns and homing endonucleases. *Fungal Genet Biol.* **46**(11), 837–48 (2009).
55. Hausner, G. Introns, mobile elements, and plasmids in *Organelle genetics - evolution of organelle genomes and gene expression*. (ed. Bullerwell, C. E.) 329–358 (Springer-Verlag, 2012).
56. Hepburn, N. J., Schmidt, D. W. & Mower, J. P. Loss of two introns from the *Magnolia tripetala* mitochondrial cox2 gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss. *Mol Biol Evol.* **29**(10), 3111–20 (2012).
57. Guha, T.K., Wai, A., Mullineux, S. T. & Hausner, G. The intron landscape of the mtDNA cytb gene among the Ascomycota: introns and intron-encoded open reading frames. *Mitochondrial DNA A DNA Mapp Seq Anal.* **29**(7), 1015–24 (2018).
58. Rudan, M. *et al.* Normal mitochondrial function in *Saccharomyces cerevisiae* has become dependent on inefficient splicing. *eLife* **7**, e35330, <https://doi.org/10.7554/eLife.35330> (2018).
59. Belfort, M. Mobile self-splicing introns and inteins as environmental sensors. *Curr Opin Microbiol.* **38**, 51–58 (2017).
60. Haugen, P. & Bhattacharya, D. The spread of LAGLIDADG homing endonuclease genes in rDNA. *Nucleic Acids Res.* **32**(6), 2049–57 (2004).
61. Dabbagh, N., Bennett, M. S., Triemer, R. E. & Preisfeld, A. Chloroplast genome expansion by intron multiplication in the basal psychrophilic euglenoid *Eutreptiella pomquetensis*. *PeerJ* **5**, e3725, <https://doi.org/10.7717/peerj.3725> (2017).
62. Martínez-Rodríguez, L., García-Rodríguez, F. M., Molina-Sánchez, M. D., Toro, N. & Martínez-Abarca, F. Insights into the strategies used by related group II introns to adapt successfully for the colonisation of a bacterial genome. *RNA Biol.* **11**(8), 1061–71 (2014).
63. Delahodde, A. *et al.* Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. *Cell.* **56**(3), 431–41 (1989).
64. Bonocora, R. P. & Shub, D. A. A likely pathway for formation of mobile group I introns. *Curr Biol.* **19**(3), 223–8 (2009).
65. Zeng, Q., Bonocora, R. P. & Shub, D. A. A free-standing homing endonuclease targets an intron insertion site in the psbA gene of cyanophages. *Curr Biol.* **19**(3), 218–22 (2009).
66. Edgell, D. R. Selfish DNA: homing endonucleases find a home. *Curr Biol.* **19**(3), R115–7, <https://doi.org/10.1016/j.cub.2008.12.019> (2009).
67. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Die Naturwissenschaften.* **58**, 465–523 (1971).
68. Szostak, N., Wasik, S. & Blazewicz, J. Hypercycle. *PLoS Comput Biol.* **12**(4), e1004853, <https://doi.org/10.1371/journal.pcbi.1004853> (2016).
69. Wolters, J. F., Chiu, K. & Fiumera, H. L. Population structure of mitochondrial genomes in *Saccharomyces cerevisiae*. *BMC Genomics.* **16**, 451, <https://doi.org/10.1186/s12864-015-1664-4> (2015).
70. Xiao, S., Nguyen, D. T., Wu, B. & Hao, W. Genetic Drift and Indel Mutation in the Evolution of Yeast Mitochondrial Genome Size. *Genome Biol Evol.* **9**(11), 3088–3099 (2017).
71. de Beer, Z. W., Marincowitz, S., Duong, T. A. & Wingfield, M. J. *Bretziella*, a new genus to accommodate the oak wilt fungus, *Ceratocystis fagacearum* (Microascales, Ascomycota). *MycKeys* **27**, 1–19, <https://doi.org/10.3897/mycokeys.27.20657> (2017).
72. Hausner, G., Reid, J. & Klassen, G. R. Do galeate-ascospore members of the Cephalosporaceae, Endomycetaceae and Ophiostomataceae share a common phylogeny? *Mycologia.* **84**, 870–881 (1992).
73. Andrews, S. FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinformatics.* <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (2010).
74. Coil, D., Jospin, G. & Darling, A. E. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics.* **31**(4), 587–9 (2015).
75. Tritt, A., Eisen, J. A., Facciotti, M. T. & Darling, A. E. An integrated pipeline for de novo assembly of microbial genomes. *PLoS One.* **7**(9), e42304, <https://doi.org/10.1371/journal.pone.0042304> (2012).
76. Beck, N. & Lang, B. F. MFannot, organelle genome annotation webserver. <http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl> (2010).
77. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**(5), 955–64 (1997).
78. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* **30**(4), 772–80 (2013).
79. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics.* **16**(10), 944–945 (2000).

80. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
81. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a largemodel space. *Syst Biol.* **61**(3), 539–42, <https://doi.org/10.1093/sysbio/sys029> (2012).
82. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One.* **5**(6), e11147, <https://doi.org/10.1371/journal.pone.0011147> (2010).
83. Gogarten, J. P. & Hilario, E. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol Biol.* **6**, 94 (2006).

Acknowledgements

GH would like to acknowledge funding from an NSERC Discovery grant. In addition, AW and AZ would like to acknowledge funding from the University of Manitoba, Faculty of Graduate Studies Graduate Enhancement of Tri-Council Stipends program. Valuable advice and assistance by members of the McGill University and Génome Québec Innovation Centre were greatly appreciated. We also would like to express our gratitude to the anonymous reviewers for their helpful suggestions.

Author Contributions

A.Z. generated the sequence data, A.Z. and A.W. analyzed the data, A.Z. prepared all figures and tables. A.Z. and G.H. designed the experiments and prepared the manuscript. A.W. provided assistance with regards to editing the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-35926-y>.

Competing Interests: The authors declare no competing interests.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018