Fungal Genetics and Biology xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Fungal Genetics and Biology

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

Comparison of mitochondrial genomes provides insights into intron dynamics and evolution in the caterpillar fungus *Cordyceps militaris*

Yongjie Zhang [a,*], Shu Zhang [a], Guozhen Zhang [b], Xingzhong Liu [c], Chengshu Wang [d], Jianping Xu [e,*]

[a] School of Life Sciences, Shanxi University, Taiyuan 030006, China
[b] Department of Plant Pathology, China Agricultural University, Beijing 100193, China
[c] State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China
[d] Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China
[e] Department of Biology, McMaster University, Hamilton, Ontario L8S 4K1, Canada

ARTICLE INFO

Article history:
Received 13 February 2015
Revised 9 April 2015
Accepted 11 April 2015
Available online xxxx

Keywords:
Cordyceps militaris
Mitochondrial genome
Intron dynamics
Horizontal gene transfer
Co-conversion

ABSTRACT

Intra-specific comparison of mitochondrial genomes can help elucidate the evolution of a species, however it has not been performed for hypocrealean fungi that form diverse symbiotic associations with other organisms. In this study, comparative analyses of three completely sequenced mitochondrial genomes of a hypocrealean fungus, *Cordyceps militaris*, the type species of *Cordyceps* genus, revealed that the introns were the main contributors to mitochondrial genome size variations among strains. Mitochondrial genes in *C. militaris* have been invaded by group I introns in at least eight positions. PCR assays of various *C. militaris* isolates showed abundant variations of intron presence/absence among strains at seven of the eight intronic loci. Although the ancestral intron pattern was inferred to contain all eight introns, loss and/or gain events occurred for seven of the eight introns. These introns invaded the *C. militaris* mitochondrial genome probably by horizontal transfer from other fungi, and intron insertions into intronless genes in *C. militaris* were accompanied by co-conversions of upstream exon sequences especially for those introns targeting protein-coding genes. We also detected phylogenetic congruence between the intron and exon trees at each individual locus, consistent with the ancestral mitochondria of *C. militaris* as having all eight introns. This study helps to explain the evolution of *C. militaris* mitochondrial genomes and will facilitate population genetic studies of this medicinally important fungus.

1. Introduction

Mitochondria play various essential roles in eukaryotic cells, particularly with respect to their primary functions in respiratory metabolism and energy production. With the emergence of next-generation sequencing in recent years, access to whole genome sequences has become easier and more affordable than ever before. Accordingly, the number of completely sequenced mitochondrial genomes (mitogenomes) of fungi has increased dramatically. To date, mitogenomes of more than 180 fungal species have been released in the public database (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=4751&opt=organelle), including 17 species within the order Hypocreales which contains some important entomogenous fungi and plant pathogens. Sizes of mitogenomes of these hypocrealean fungi vary between 22.4 kb and 95.7 kb (data obtained from public database), but they generally contain 14 conserved mitochondrial protein-coding genes, two rRNA genes and a variable number of tRNA genes with almost identical gene order (Eldarov et al., 2015). Their genome size variations are primarily due to differences in the number of introns and the length at inter-genic regions.

Intra-specific comparative analysis of mitogenomes can help elucidate the evolution of a species and identify highly polymorphic regions and/or strain-specific markers (Bartelli et al., 2013; Formey et al., 2012). So far, intra-specific comparison of whole mitogenomes have been performed in seven fungal species: *Candida albicans* (Saccharomycetales, Saccharomycetes, Ascomycota) (Bartelli et al., 2013), *Lachancea kluyveri* (Saccharomycetales, Saccharomycetes, Ascomycota) (Jung et al., 2012), *Lachancea thermotolerans* (Freel et al., 2014), *Mycosphaerella graminicola* (Capnodiales, Dothideomycetes, Ascomycota) (Torriani et al., 2008), *Neurospora crassa* (Sordariales, Sordariomycetes, Ascomycota) (McCluskey, 2012), *Podospora anserina* (Sordariales, Sordariomycetes, Ascomycota) (Cummings et al., 1990), and

* Corresponding authors. Tel.: +1 905 525 9140 27934 (J. Xu).
E-mail addresses: zhangyj2008@sxu.edu.cn (Y. Zhang), jpxu@mcmaster.ca (J. Xu).

http://dx.doi.org/10.1016/j.fgb.2015.04.009
1087-1845/

Rhizophagus irregularis (Glomerales, Glomeromycetes, Glomeromycota) (Formey et al., 2012). Because species of Hypocreales form diverse symbiotic associations that include antagonistic to mutualistic interactions with numerous animals, plants, and other fungi (Sung et al., 2008), it is interesting to understand the mitochondrial DNA (mtDNA) evolution of a hypocrealean species. Variations of mtDNA among different isolates of the same hypocrealean species, however, are currently unknown. Although there are mitogenome data from two *Beauvera bassiana* isolates in the public database (Ghikas et al., 2010; Xu et al., 2009a), they likely represent two different species (see Results of this study).

Cordyceps militaris, which generally parasitizes larva or pupa of lepidopteran insects, is the type species of the genus *Cordyceps* (Hypocreales, Ascomycetes). This fungus is distributed worldwide from 0 to >2000 m above sea level (Shrestha et al., 2012). Its fruiting body has now been mass-produced artificially and developed into health food (i.e. not just nutritious in an ordinary sense, but eaten specifically for its health-promoting properties), and the species is one of the most representative and widely-used species in *Cordyceps* sensu lato (Shrestha et al., 2012). Biologically active compounds (e.g., cordycepin, polysaccharides, cordymin) isolated from the fungus exhibit a variety of pharmacological effects, including anti-cancer, antioxidant, anti-inflammatory, immune-enhancing, or antifungal activities (Das et al., 2010; Tuli et al., 2013; Wang et al., 2013; Wong et al., 2011). In addition to the interest in artificial cultivation and pharmacological effects, researchers have studied this fungus broadly from the viewpoints of genomics (Zheng et al., 2011), transcriptomics (Xiong et al., 2010; Yin et al., 2012), and proteomics (Yin et al., 2012). Investigation on intra-specific genetic variation in *C. militaris*, however, has been rather limited (Zhang et al., 2013).

During the preparation of this paper, the mitogenome of one *C. militaris* strain, EFCC-C2, was submitted to the public database (Sung, 2015). Therefore, the objectives for this work were: (1) to perform the first comparative mitogenomic analysis of hypocrealean fungi in *C. militaris*, (2) to document intron presence/absence dynamics in mitochondrial genes in *C. militaris*, and (3) to characterize the ancestral intron distribution pattern and the evolution of introns in *C. militaris*. In this study, we reported the complete mitogenomes of two new *C. militaris* isolates, CM01 and V26-17. Intra-specific comparison of the two isolates plus EFCC-C2 revealed intron presence/absence dynamics which were further expanded by investigating additional isolates. The ancestral state of *C. militaris* was inferred to have introns at all eight investigated loci. Intron loss was the dominant event; however, independent intron gain after intron loss was also inferred for at least two introns. All mitochondrial introns of *C. militaris* were acquired probably through horizontal gene transfer likely at the beginning of *C. militaris* speciation, and intron invasions were often accompanied by co-conversions of upstream exon sequences flanking intron insertion sites. This study broadens our understanding of the mtDNA evolution of hypocrealean fungi.

2. Materials and methods

2.1. Fungal materials

Cordyceps militaris strains CM01, whose nuclear genome was previously reported (Zheng et al., 2011), and V26-17, a mono-ascospore isolate, were employed to obtain full-length mitogenome sequences. The two isolates plus the *C. militaris* isolate EFCC-2C, whose mitogenome was recently reported (Sung, 2015), were used to perform comparative mitogenomics. Isolates CM01, V26-17, and 18 other *C. militaris* isolates (Supplementary Table 1) were used in PCR assays to investigate presence/absence dynamics of introns and other sequences. The 20 isolates showed only one nucleotide difference in the 528 bp of the internal transcribed spacers of the nuclear ribosomal rRNA gene cluster, consistent with these strains belonging to the same species. To obtain their genomic DNA, these isolates were cultivated at 25 °C for 10 days in PDA media with cellophane paper covering the medium surface. Mycelia were collected and used to extract total genomic DNA using the CTAB method (Zhang et al., 2010).

2.2. Assembly, annotation, and comparative analysis of different mitogenomes of C. militaris

BLAST analysis against the *C. militaris* CM01 genome data (NCBI accession no. AEVU00000000), using *Beauveria bassiana* mitogenome (NC_017842) as the query, revealed one scaffold (i.e., CCM_S00011) harboring mitochondrial genes, but this scaffold was incomplete due to existence of nine gap regions. To fill the gaps and complete the scaffold, primers were designed based on known sequences of CCM_S00011. PCRs were performed using the DNA polymerase KOD FX (TOYOBO Bio-Technology Co. LTD, Japan), and sequences of amplicons were determined by Sanger sequencing at SinoGenoMax Co. Ltd. (Beijing, China) or at the Mobix lab of McMaster University (Hamilton, Canada). To obtain the complete mitogenome of V26-17, multiple overlapping PCR primer sets were designed based on the assembled mtDNA sequence of CM01, and amplicons were sequenced and assembled. All primers used in this study were designed using the online software Primer3 (Untergasser et al., 2012); primer information is available upon request.

The mitogenome of *C. militaris* was first annotated automatically using the MFannot tool (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl) based on genetic code 4. The MFannot program can predict protein-encoding genes, rRNAs, tRNAs, and intron types. The predictions were individually checked as follows. First, the boundaries of rRNAs (especially for the 5′ and 3′ termini and exons of *rnl*) were determined by aligning with other hypocrealean fungi whose mitogenomes have been reported. Second, transfer-RNA (tRNA) annotations were further confirmed by tRNAscan-SE 1.21 (Schattner et al., 2005). Third, intron–exon boundaries were confirmed by aligning corresponding sequences with an intron-less gene from a related species or with other *C. militaris* isolates without corresponding introns. Fourth, intronic and inter-genic regions were further searched by ORF Finder (http://www.ncbi.nlm.nih.gov) to identify additional ORFs (open reading frames) larger than 300 bp. Functional assignments of these new ORFs were made based on sequence similarity to characterized fungal mitochondrial proteins using BLASTP searches against NCBI databases (http://www.ncbi.nlm.nih.gov) and HMM searches against the protein families in the Pfam 27.0 database (Finn et al., 2014). ORFs that had no significant similarity to known genes and were at least 100 amino acids in length were annotated as hypothetical proteins.

The mitogenomes of three *C. militaris* isolates, CM01, V26-17, and EFCC-C2, were aligned using Mauve 2.3.1 (Darling et al., 2010) to illustrate positions of introns and major alignment gaps. Alignments of each exonic or intronic region were made by Muscle (Edgar, 2004) as implemented in MEGA 6.0 (Tamura et al., 2013).

2.3. Phylogenetic analysis of hypocrealean species

To determine the evolutionary relationships between *C. militaris* and other hypocrealean species, concatenated amino acid sequences of *atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, and *nad6*, a total of 4,195 characters, were used in a phylogenetic analysis for all hypocrealean species whose

Please cite this article in press as: Zhang, Y., et al. Comparison of mitochondrial genomes provides insights into intron dynamics and evolution in the caterpillar fungus *Cordyceps militaris*. Fungal Genet. Biol. (2015), http://dx.doi.org/10.1016/j.fgb.2015.04.009