Dynamic Evolution of Eukaryotic Mitochondrial and Nuclear Genomes: A Case Study in the Gourmet Pine Mushroom Tricholoma matsutake

Zhang, Shu

2021-11

Zhang , S , Bai , X , Ren , L-Y , Sun , H-H , Tang , H-P , Vaario , L-M , Xu , J & Zhang , Y-J 2021 , ' Dynamic Evolution of Eukaryotic Mitochondrial and Nuclear Genomes: A Case Study in the Gourmet Pine Mushroom Tricholoma matsutake ' , Environmental Microbiology , vol. 23 , no. 11 , pp. 7214-7230 . https://doi.org/10.1111/1462-2920.15792

http://hdl.handle.net/10138/349471
https://doi.org/10.1111/1462-2920.15792

acceptedVersion

Xu Jianping (Orcid ID: 0000-0003-2915-2780)
Zhang Yongjie (Orcid ID: 0000-0002-6331-2189)

Dynamic Evolution of Eukaryotic Mitochondrial and Nuclear Genomes: A Case Study in the Gourmet Pine Mushroom *Tricholoma matsutake*

Shu Zhang[1], Xue Bai[1], Li-Yuan Ren[1], Hui-Hui Sun[1], Hui-Ping Tang[1], Lu-Min Vaario[2*], Jianping Xu[3*], Yong-Jie Zhang[1*]

[1] School of Life Science, Shanxi University, Taiyuan 030006, China;

[2] Department of Forest Science, University of Helsinki, Helsinki FI-00014, Finland;

[3] Department of Biology, McMaster University, Hamilton, Ontario, L8S 4K1, Canada

***Corresponding authors:**

Lu-Min Vaario Department of Forest Science, University of Helsinki, Helsinki FI-00014, Finland; e-mail: lu-min.vaario@helsinki.fi

Jianping Xu Department of Biology, McMaster University, Hamilton, Ontario, L8S 4K1, Canada; Phone +1 905 525 9140-27934; email jpxu@mcmaster.ca

Yong-Jie Zhang School of Life Sciences, Shanxi University, Taiyuan 030006, China; Telephone: +86 135 4638 0352, email: zhangyj2008@sxu.edu.cn

Running title: Dynamic genome evolution of the pine mushroom

Conflict of Interest: The authors declare no conflict of interest.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.15792

Originality-Significance Statement

1. This study performed the first comparative mitogenomic study and greatly enriched the available genomic information of *Tricholoma matsutake* by performing high-throughput DNA sequencing of individuals originating from various regions.

2. Comparison on different mitogenomes revealed low intraspecific variations in *Tricholoma matsutake* except for one *cox1* intron cox1P372 that showed presence/absence dynamics among different individuals.

3. There was clear evidence of transfer of mitochondrial DNA into the nuclear genome. However, mitochondrial DNA of *T. matsutake* showed a lower mutation rate than nuclear DNA, and topologies of phylogenetic trees derived from mitogenome and nuclear DNA sequences were incongruent. Together, this study revealed the dynamic genome evolution of the gourmet pine mushroom.

Summary

Fungi, as eukaryotic organisms, contain two genomes, the mitochondrial genome and the nuclear genome, in their cells. How the two genomes evolve and correlate to each other is debated. Herein, taking the gourmet pine mushroom *Tricholoma matsutake* as an example, we performed comparative mitogenomic analysis using samples collected from diverse locations and compared the evolution of the two genomes. The *T. matsutake* mitogenome encodes 49 genes and is rich of repetitive and non-coding DNAs. Six genes were invaded by up to 11 group I introns, with one *cox1* intron cox1P372 showing presence/absence dynamics among different samples. Bioinformatic analyses suggested limited or no evidence of mitochondrial heteroplasmy. Interestingly, hundreds of mitochondrial DNA fragments were found in the nuclear genome, with several larger than 500 nt confirmed by PCR assays and read count comparisons, indicating clear evidence of transfer of mitochondrial DNA into the nuclear genome. Nuclear DNA of *T. matsutake* showed a higher mutation rate than mitochondrial DNA. Furthermore, we found evidence of incongruence between phylogenetic trees derived from mitogenome and nuclear DNA sequences. Together, our results reveal the dynamic genome evolution of the gourmet pine mushroom.

Keywords: *Tricholoma matsutake*, mitogenome, nuclear genome, phylogeny, evolution, comparative mitogenomics

Introduction

Mitochondria are essential organelles present in most eukaryotic cells. They ensure the supply of energy in the form of ATP that can be used for cellular processes. In addition, mitochondria are also integral to the regulation of metabolism, apoptosis, and cytosolic calcium concentration (Chatre and Ricchetti, 2014). Mitochondria contain their own genetic materials — a vestigial genome originating from an endosymbiotic α-proteobacterial ancestor. Mitochondrial functions depend on the expression of both nuclear-encoded genes and mitochondria-encoded genes (Woodson and Chory, 2008). More than 1000 mitochondrially located proteins are encoded in the nuclear genome, and the latter controls essential processes needed for proper functioning of mitochondria, including mitochondrial DNA (mtDNA) maintenance, mitochondrial protein synthesis, coenzyme Q10 biosynthesis and assembly of respiratory chain complexes (Woodson and Chory, 2008; Isaac *et al.*, 2018). Mitochondrial genomes (mitogenomes) are distinct from nuclear genomes in terms of architecture, size and complexity, copy number per cell, mutation rate, inheritance, replication and partition patterns during cell division and so forth (Xu and Li, 2015; Xu and Wang, 2015; Sandor *et al.*, 2018). It is, therefore, important to understand the mitogenome structures and evolution.

As a large group of eukaryotic organisms, fungi contain mitogenomes in their cells. Over recent years, mitogenomes of an increasing number of fungal species are sequenced. For example, at the end of 2020, mitogenomes from over 600 fungal species have become available on public nucleotide databases. These mitogenomes

show large variations in size and structure both among individuals of a given species as well as among fungal species (Zhang *et al.*, 2015; Zhang *et al.*, 2017b; Zhang *et al.*, 2017a; Zhang *et al.*, 2017e; Zhang *et al.*, 2017c; Fan *et al.*, 2019; Nie *et al.*, 2019; Zhang *et al.*, 2019a).

Tricholoma matsutake (S. Ito & S. Imai) Singer is an ectomycorrhizal gourmet mushroom in the phylum Basidiomycota. It forms symbiotic relationships with the roots of trees belonging to families Pinaceae and Fagaceae in the Northern Hemisphere (Vaario *et al.*, 2017; Wang *et al.*, 2017a; Yamanaka *et al.*, 2020). The fruiting body of the fungus, which is predominantly produced during autumn, is of great commercial value in East Asia (Wang *et al.*, 2017a), and has shown notable biological activities and medicinal values (Yue *et al.*, 2017). There have been significant declines of the annual global harvest, while attempts to artificially cultivate fruiting bodies of *T. matsutake* have been unsuccessful (Wang *et al.*, 2017a). Indeed, *T. matsutake* has been categorized as Vulnerable on the IUCN (International Union for Conservation of Nature) Red List of Threatened Species (Brandrud, 2020) and is classified as a protected species in China (Winkler, 2008).

Effective conservation of natural resources and establishment of suitable management strategies require a comprehensive understanding of the population structure of *T. matsutake*. Genetic diversity and population genetics of the fungus have been investigated using various molecular markers, such as SSR (Zeng and Chen, 2015; Kurokochi, 2021), PCR-RFLP (Xu *et al.*, 2008), and single-nucleotide polymorphic (SNP) DNA markers (Sandor *et al.*, 2020). These studies mainly used

samples from restricted regions, and most of them revealed genetic differentiation among different geographical populations.

To date, mitogenome sequences of three *T. matsutake* strains have been recorded in GenBank. In the paper first publishing the *T. matsutake* mitogenome (i.e., JX985789) in 2016, annotations were incomplete because introns, intronic and intergenic ORFs (open reading frames) were not described at all (Yoon *et al.*, 2016). In a more recent paper, the mitogenome of another *T. matsutake* strain (i.e., MN873034) was thoroughly annotated (Huang *et al.*, 2021), but it was smaller than JX985789 (74865 bp vs. 76037 bp). A third mitogenome of *T. matsutake* (LC385608, 76089 bp) with a size similar to JX985789 is also present in GenBank, but it has not been officially described in a peer-reviewed publication. Indeed, the extent of intraspecific mitogenome divergence among *T. matsutake* samples remains largely unknown. We recently examined the phylogeography of the global matsutake species complex (involving *T. matsutake* and three other related species) using sequences at the ITS region and four mitochondrial loci, and found that the ITS locus had a higher level of polymorphism than mitochondrial loci (Sandor *et al.*, 2020). However, sequence divergence at the ITS locus is not necessarily indicative of sequence divergence in the rest of the nuclear genome. A comprehensive understanding of the evolution of *T. matsutake* requires the analyses of both the mitochondrial and nuclear DNA sequences at multiple genes.

Due to its wide distribution across two continents, we hypothesize that there are high intraspecific variations among *T. matsutake* samples from distantly distributed

localities. Further, due to their inherent differences, we hypothesize that mitochondrial and nuclear DNAs may show differences in their evolution rates and phylogenetic topologies. To test these hypotheses, we assembled mitogenomes of *T. matsutake* samples collected from distant localities, and compared them with their nuclear DNAs. Our specific goals were (1) to provide a detailed annotation of the *T. matsutake* mitogenome, (2) to reveal intraspecific variations among different *T. matsutake* mitogenomes, (3) to investigate if there is frequent transfer of mitochondrial DNA into the nuclear genome, and (4) to compare the evolution between mitochondrial and nuclear DNAs.

Results

Organization of the *T. matsutake* mitogenome

Two programs NOVOPlasty and GetOrganelle were used for assembling mitogenomes from high-throughput sequencing reads generated in this study. Both programs succeeded in generating circular mitogenomes from 12 out of the 17 samples used in this study, and sequences resulting from both programs were identical. For four other samples (JT5, T8, SU3, and TM-NH), both programs assembled into 1–3 large incomplete mitochondrial contigs. Gaps between these contigs or between both ends of a single contig were later filled and completed successfully by PCR and Sanger sequencing. For the sample Tm_Korea, both programs generated multiple short incomplete mitochondrial contigs, and thus this sample was excluded from downstream analyses.

Since the first paper describing the *T. matsutake* mitogenome lacked several

important pieces of annotation information (Yoon *et al.*, 2016), here we provide a full description of the *T. matsutake* mitogenome using NF5 as a representative. This NF5 strain, which originated from a Scots pine dominated forest located in the Arctic Circle in Finland, represented the most dominant mitogenome-type and one of the largest mitogenomes of *T. matsutake* (see later section), and had a mitogenome size close to the published mitogenome JX985789 (Table 1).

The NF5 mitogenome was a circular molecule of 76,034 bp with an AT content of 79.31%. This mitogenome encoded a total of 49 genes (excluding intronic ORFs), including two ribosomal RNAs (*rnl* and *rns*), 26 tRNAs, 14 conserved proteins of the oxidative phosphorylation system (*nad1-6*, *4L*; *cob*; *cox1-3*, and *atp6*, *8*, *9*), and seven intergenic ORFs (Fig. 1A; Supporting Table S1). These tRNA genes covered all 20 standard amino acids. The seven intergenic ORFs encoded either ribosomal protein S3 or hypothetical proteins. Most genes were transcribed at the forward strand, but 10 genes (*atp6*, *atp8, orf109*, *orf132A*, and 6 tRNA genes) were transcribed at the reverse strand. For the two neighboring gene pairs commonly found in other fungi, *nad3* immediately followed *nad2*, while *nad5* overlapped one nucleotide with its upstream gene *nad4L*. All other neighboring genes were separated by 24–3650 bases. Intergenic regions (i.e., those regions among the above-mentioned 49 genes) accounted for 43.7% (33,215 bp) of the whole mitogenome (Fig. 1B), and 13 intergenic regions were larger than 1000 bp. The largest interval (3650 bp) was seen between *orf335* and *rnl*.

A total of 11 introns were found within six genes, including *cob* (1 intron), *cox1* (5), *cox2* (2), *nad1* (1), *nad5* (1), and *rnl* (1) (Table 2). These introns all belonged to

the group I intron family but fell into three specific subgroups, namely IA (1 intron), IB (9), and ID (1). Except for four introns (cobP823, cox2P318, nad1P276, and nad5P417), all other introns contained putative ORFs encoding for LAGLIDADG homing endonucleases. The four ORF-lacking introns as well as the ORF-containing intron cox1P372 might be degenerating, as evidenced by the detection of frame shifts and/or nonsense mutations. Intronic regions (14,741 nt in total, including intronic ORFs) accounted for 19.4% of the mitogenome, or 34.4% of genic regions.

We observed no large-fragment duplication within any of the mitogenomes. According to BLAST results of the mitogenome against itself, only 13 short-fragment duplication events were identified at alignment lengths of 29–166 nt with nucleotide identities of 89.4–100%. The 166-nt alignment resulted from the duplication of *trnI* (Supporting Table S2). More than 150 tandem repeats were found, with a widespread distribution across the whole mitogenome, including several ones whose total lengths were larger than 100 nt (Supporting Table S3). Most tandem repeats occurred at intergenic (77.4%) or intronic (15.5%) regions, and few occurred at exonic regions (7.1%). A total of 767 SSRs were identified (Supporting Table S4), with the most abundant motifs being pentamers (311, 41%) and hexamers (240, 31%). These SSR motifs repeated mostly twice (646, 84%), followed by three times (61, 8%). The length distribution of these SSRs were mainly 10 bp (306, 40%) and 12 bp (269, 35%), but nine SSRs larger than 30 bp (5 pentamers, 3 hexamers, and 1 heptamer) were also found. Most SSRs identified in the mitogenome were located at intergenic regions (69.0%), followed by exonic regions (17.7%) and intronic regions (13.3%). In

addition, according to REPuter analysis results, there were 96 reverse, 75 forward, 31 palindromic, and 29 complement repeats in the mitogenome, at a length range of 24–159 nt (Supporting Table S5). The 159-nt repeat corresponded to a forward repeat event targeting the two copies of *trnI* plus their both flanking regions.

Origin of introns in the *T. matsutake* mitogenome

As seen from the above analysis of repetitive elements and based on sequence comparisons among introns, the possibility of intron origin through intra-genomic duplication events could be excluded. We therefore performed online BLASTN searches of *T. matsutake* introns against the public nucleotide database (Supporting Table S6). All introns except nad1P276 showed significant hits to introns present in other fungi, and target introns were always found in orthologous genes. This supported the origin of introns through horizontal gene transfer.

Intra-specific variation of *T. matsutake* mitogenomes

Intra-specific comparison of *T. matsutake* mitogenomes were performed using the 16 mitogenomes assembled in this study and the three mitogenomes downloaded from GenBank. These 19 mitogenomes varied from 74,750 bp to 76,089 bp, with 14 having a bigger mitogenome of approx. 76.1 kb and 5 having a smaller mitogenome of approx. 74.8 kb (Table 1). This mitogenome size variation did not directly correlate with countries where these samples originated from (Fig. 2A), and with host species or elevation neither. Those with bigger mitogenomes included all samples from Japan (3), South Korea (1) and Sweden (3), five from Finland, and two from Northeast China. Those with smaller mitogenomes included three samples from Southwest

China, and one each from Finland and Canada. It should be noted that samples from Southwest China and those from Northeast China differed on mitogenome sizes.

Size variations among the bigger and smaller mitogenomes mainly resulted from the presence or absence of one *cox1* intron (namely cox1P372). This 1290-bp intron was present in all samples with bigger mitogenomes (and also in Tm_Korea with incomplete mitogenome) and absent in all samples with smaller mitogenomes (Figs. 2A & 3). According to the inference of ancestral intron state, the root node of *T. matsutake* was suggested to have cox1P372 with an occurrence of 99.7%, and there seemed to be only one loss event of cox1P372 (Fig. 2B). It should be noted that this intron, when present, may no longer be mobile due to stop codon mutations in its homing endonuclease gene. Interestingly, we observed no co-conversion of *cox1* exon sequences flanking the intron insertion site between intron-containing samples and intron-lacking samples. Indeed, there was no sequence variation at cox1P372 or the whole exon sequences of *cox1*, consistent with their recent shared ancestries. The remaining introns were all present in all samples.

Analyses of whole mitogenome sequence variations grouped these 19 strains into 11 different mitogenome sequence types (designated mito-types A to K; Fig. 2A). The mito-type A was the most dominant one and was shared by eight samples from Finland and Sweden. The mito-type H was shared by two samples, one each from Finland and China. Other nine mito-types were each represented by one sample. Mito-types H, I, J and K all lacked the *cox1* intron cox1P372, whereas other mito-types all contained the intron. Mito-types D and E found in isolates F2 and W2

respectively were completely identical except for 1-bp difference of polyT at the *atp6*/*atp9* intergenic region.

Gene contents and order among different *T. matsutake* samples were highly conserved except for few differences resulting from nucleotide variations (Supporting Table S7). Consistent to this, the overall nucleotide variations of mitogenomes among the 19 samples were low. Among the 49 genes detected in *T. matsutake* mitogenomes, only 16 showed nucleotide variations, and others were invariable (Supporting Table S8). Alignment of the 19 mitogenomes generated totally 75,081 characters in length after excluding cox1P372. Of these characters, 74,451 were invariable, 510 were sites with alignment gaps, and 120 were variable (polymorphic) sites (Table 3). Of the 120 variable sites, 68 were singleton mutations, and 52 were parsimony informative sites that were shared by at least two strains. This corresponded to a genetic divergence of 0.84% (SNP + indels) and a SNP frequency of 0.16%. Among the different parts of the mitogenome, the intergenic regions (1.31%/0.25%, genetic divergence/SNP frequency) showed a higher variation than genic regions (0.46%/0.09%). For intron-containing genes, intronic regions (0.25%/0.12%) generally showed a higher variation than exonic regions (0.14%/0.08%); however, there was an opposite trend for *rnl* and protein-encoding genes. For *rnl*, exons (0.23%/0.11%) were more divergent than introns (no variation), while for protein-encoding genes, introns (0.27%/0.13%, on average) were more divergent than exons (0.06%/0.06%, on average).

We defined the 14 samples with larger mitogenome as one group (designated as

Pop1) and the 5 samples with smaller mitogenomes as the second group (designated as Pop2). Pop1 had 85 polymorphic sites, including 54 singleton variable sites and 31 parsimony informative sites. Pop2 showed 25 polymorphic sites, including 19 singleton variable sites and 6 parsimony informative sites. For the 120 polymorphic sites detected among all samples, three were mutations shared by Pop1 and Pop2, 82 were polymorphic in Pop1 but monomorphic in Pop2, 22 were polymorphic in Pop2 but monomorphic in Pop1, and 13 were fixed differences between Pop1 and Pop2 (i.e., nucleotide sites at which all of the sequences in one group are different from all of the sequences in the second group). For the 13 fixed sites, 8 located at intergenic regions, 4 at exonic regions (*rnl*-E1, *cob*-E1, *cox2*-E3, and *nad5*-E2), and 1 at an intronic region (Supporting Table S9).

Detection of heteroplasmy in *T. matsutake* mitogenomes

We examined potential heteroplasmy (i.e., the occurrence of different types of mitochondrial alleles in the same cell) for the 16 samples whose complete mitogenomes were successfully assembled in the present study. Two samples (JT5 and TM-NH) did not show any heteroplasmy at the defined allele frequency threshold of 3%, while 1–7 putative heteroplasmic positions were suggested for each of other samples (Table 4). The frequency of alternative bases at these positions was generally smaller than 10% except two sites in Tm1 whose frequencies were at 15–20%. Although the frequency of the two alternate sites of Tm1 reached the detection limit of Sanger sequencing (10–20%), we cannot examine the two sites by PCR due to the unavailability of genomic DNA of this sample. Nevertheless, we amplified regions

around several putative heteroplasmic sites from three strains EF, T8, and JT89, and no detectable heterozygosity was found by checking their sequencing chromatograms, consistent with the low frequency of the alternative alleles in these individuals (Supporting Fig. S1).

Detection of NUMTs in *T. matsutake* nuclear genomes

BLAST analyses of mitogenomes of NF5, F2, and L1 against their corresponding nuclear genomes generated 262–306 hits at alignment lengths of 32–979 nt and identities of 80–100% (Supporting Table S10). LAST analyses suggested a small number of hits (160–222 hits per sample) after filtration (Supporting Table S11). The three samples showed similar but different NUMT frequencies (Supporting Fig. S2A). Taking the 262 NUMTs of NF5 detected by BLAST as an example, most of them were shorter than 100 nt (190 in number, 72.5%), and few were 100–500 nt (68, 26.0%) or > 500 nt (4, 1.5%). The mitochondrial fragments related to these NUMTs were scattered across the mitogenome (Supporting Fig. S2A), suggesting that there was no clear hot spots for the transfer of mitochondrial DNA into the nucleus.

We focused on those hits with alignment lengths > 500 nt. Each of the three *T. matsutake* samples under survey (i.e., NF5, F2, and L1) had 4–5 such large hits (Supporting Table S12; Supporting Fig. S2A). They were homologous to mitochondrial fragments *rns* (designated as NUMT1), *orf109* & both flanking regions (NUMT2), *nad1*-E2 (NUMT3), *cox2/cox3* intergenic region (NUMT4), and *cox2*-i2 (cox2P357, NUMT5). PCR surveys and Sanger sequencing confirmed all these large NUMTs representing authentic sequences (Supporting Fig. S2B). Read counts of

these NUMTs were obviously lower than mitogenomes (Supporting Fig. S3), which supported that these NUMTs were located in nuclear genomes. All these large NUMTs except NUMT2 might exist as a single copy as deduced from their presence in only one nuclear contig at least in the three investigated samples (Supporting Table S12). NUMT2 was seen in two nuclear contigs in each sample, indicating that it had duplicated since its insertion into the nuclear genome. It should be noted that the two duplications of NUMT2 (designated as NUMT2a and NUMT2b for its two copies) might have diverged since then because NUMT2a and NUMT2b only had a low similarity of 90%, slightly lower than their similarities to their mitochondrial partner (92–93%). In addition, NUMT2a/NUMT2b sequences among the three samples had a high similarity (95–99%). It was also interesting to mention that NUMT1 did not show significant match to its nuclear partner (*nrSSU*) although it had a similarity of 96% to its mitochondrial partner (*rns*). This supported its origin from mitochondrial DNA. Actually, online BLAST searches of each of these large NUMTs showed significant hits to mitochondrial fragments instead of nuclear fragments of *Tricholoma* spp., which confirmed their origins from mitogenomes.

Evolutionary comparison between mitochondrial and nuclear DNAs

We determined nucleotide variations at 12 nuclear markers among the 16 samples with complete mitogenomes assembled in this study. No nucleotide variation was observed on *act* and *nrSSU*, whereas other 10 nuclear markers all showed certain variations, though at a variable extent of divergence (Table 5). Their SNP divergences varied from 0.06% (*nrLSU*) to 1.53% (*Amy1*), and their genetic frequencies (SNP +

indels) varied from 0.06% (*nrLSU*) to 2.33% (ITS). Among the nine protein-encoding genes, *Amy1* was the most polymorphic. Among the three nrDNA fragments, the ITS region was the most polymorphic and displayed a similar level of divergence to certain protein-encoding genes. Overall, these 12 nuclear markers displayed a genetic divergence of 0.70–0.79% and a SNP frequency of 0.59–0.66%, depending on whether *act* and *nrSSU* were included for calculation or not.

According to concatenated nuclear sequences at the 12 loci, the 16 samples belonged to 12 multi-locus haplotypes, more than the 8 haplotypes deduced from mitogenomes from the same group of samples (Fig. 4). Sequences of NF5 and SW4 were identical, and sequences of C4, N18, NF6, and T8 were identical at all 12 nuclear loci. Each of the remaining samples belonged to a unique haplotype. Nuclear markers showed a higher haplotype diversity (0.94 ± 0.048) and nucleotide diversity (0.0016 ± 0.00023) than mitogenomes, which was 0.75 ± 0.012 and 0.00035 ± 0.00005, respectively.

The topologies of the tanglegram generated by the analyses of nuclear and mitochondrial markers were compared (Fig. 4). As seen from direct observations, though few samples were grouped together in both trees (e.g., C4, N18, NF6, and T8; NF5 and SW4; EF and L1), the two topologies showed inconsistency. It should be noted that the three samples, C4, N18, and T8, were from the same sampling locality but different shiros, which are mycelial aggregations that develop in association with the roots of coniferous trees and soil particles and the places where basidiocarps of *T. matsutake* occurring in groups. As expected, these three samples showed identical

mitochondrial/nuclear sequences. KH, SH, and PHT tests showed that the difference of the two topologies were statistically significant ($P \leqslant 0.001$; Supporting Table S13). Similar results were obtained when just looking at the 12 samples in Pop1; however, when just looking at the 4 samples in Pop2, they were not phylogenetically different.

Mantel test, however, showed that the mitochondrial and nuclear genetic distances among the overall 16 samples or among the 12 Pop1 samples were significantly correlated (r=0.69–0.79, $P < 0.001$) even when controlling the effect of geographical distances (Supporting Table S14). The correlation between mitochondrial/nuclear genetic distance and geographical distance were also significant, but it was not always true when the effect of a third distance not in comparison (nuclear or mitochondrial genetic distance) was controlled (Supporting Table S14).

Discussion

In this study, we provided a detailed description of the *T. matsutake* mitogenome with NF5 as the representative strain. The *T. matsutake* mitogenome encoded 49 free-standing genes (including rRNA and tRNA coding genes, core protein coding genes and intergenic ORFs, but excluding intronic ORFs). Genic regions accounted for 56.3%, and intergenic regions accounted for 43.7% of the mitogenome. For *Tricholoma* species, it seems common that their mitogenomes have a high proportion of non-coding intergenic regions (34.2–45.5%, 40.7% on average) (Huang *et al.*, 2021), while some other fungal lineages might have a low proportion of intergenic

regions. For instance, intergenic regions accounted for 13.9–33.8% (21.7% on average) in mitogenomes of Ophiocordycipitaceae (Zhang *et al.*, 2017c). There was no large fragment duplication within the *T. matsutake* mitogenome. However, we detected many tandem repeats and SSRs. These repetitive elements may be useful for genotyping *T. matsukate* samples from different geographic regions for population genetic studies. Interestingly, the nuclear genome of the fungus is also rich in non-coding DNA and repetitive elements, including transposable elements contributing to its large nuclear genome size of 189.0 Mbp (Min *et al.*, 2020). The *T. matsutake* NF5 mitogenome also contained 11 introns distributed in six genes, and these introns most likely originated from other fungi by horizontal gene transfer, which is a common phenomenon reported in many fungi (Zhang *et al.*, 2015; Zhang *et al.*, 2017c; Fan *et al.*, 2019; Zhang *et al.*, 2019a).

We performed the first comparative study of *T. matsutake* mitogenomes using samples from distantly separated localities. Besides some polymorphic and indel sites, one obvious finding was cox1P372, one of the *cox1* introns, which showed presence/absence dynamics among different *T. matsutake* samples. This intron is also absent in reported mitogenomes of four other *Tricholoma* species (Huang *et al.*, 2021). According to the presence/absence pattern of cox1P372 and nucleotide variations across the whole mitogenomes, the 19 *T. matsutake* samples fell into two different groups (Fig. 2A). There were some fixed mutations between the two groups (Supporting Table S9), but how the two groups differentiate on morphology and physiology remains unknown. The fungus seemed to contain cox1P372 at the time of

speciation, and the loss of cox1P372 seemed to be a derivative character (Fig. 2B). Loss of cox1P372 seemed to have occurred at least in three distant regions, namely Southwest China, Finland, and Canada. The three sampling regions are different in altitude, latitude, temperature, precipitation, and host species, and thus the loss of cox1P372 may not be driven by ecological factors. The cox1P372 locus may act as a potential marker to discriminate *T. matsutake* samples from different localities, especially between those from Northeast China and those from Southwest China. Samples from the two Chinese regions are sold at distinctly different prices, and there are high proportion of counterfeiting due to wrong claims of sample origin in matsutake markets (Xu *et al.*, 2010). The cox1P372 locus may also act as a potential marker to discriminate *T. matsutake* from other closely-related species because it's absent in other known mitogeonomes of *Tricholoma* species, and to facilitate the study of mitochondrial inheritance of *T. matsutake* because of its presence/absence dynamics among different *T. matsutake* samples.

Most eukaryotic cells possess a population of mitochondria, in the sense that mtDNA is held in multiple copies per cell, where the sequence of each molecule can vary. Hence, intra-cellular mitochondrial heterogeneity is possible, which can induce inter-cellular mitochondrial heterogeneity. In this study, we examined mitochondrial heteroplasmy in *T. matsutake* for the first time. Using the normal assembly function of NOVOPlasty, we did not detect any degenerate bases in all *T. matsutake* mitogenomes, and the resulting sequences were identical to those obtained by the program GetOrganelle. When assembling mitogenomes of other fungi by our research

group using NOVOPlasty, we sometimes observed some degenerate bases in assembled mitogenomes of few other fungi (Unpublished data), probably indicating mitochondrial heteroplasmy. When switching to the heteroplasmy detection function, however, NOVOPlasty reported few sites with alternative bases in several *T. matsutake* samples, though at a low frequency (Table 4). PCR assays were performed to examine some of these alternative sites, but no heterozygosity was clearly observed in sequencing chromatograms (Supporting Fig. S1). Until now, mitochondrial heteroplasmy has not been frequently reported in fungi (Wang *et al.*, 2017b). It is possible that those alternative bases detected in *T. matsutake* are caused by the existence of abundant NUMTs. It is also possible those alternative bases are transient, followed by rapid segregation of parental mtDNA genotypes into homoplasmic progeny cells. In addition, their origin from Illumina systemic errors cannot be ruled out although clean reads were used in related analyses.

Following the acquisition of mitochondria by eukaryotic cells during endosymbiotic evolution, most of the genes in mitochondria were either lost or transferred to the nucleus (Adams and Palmer, 2003). Those sequences, which originate from the invasion of nuclear DNA by mtDNA, are called NUMTs. In this study, a large number of NUMTs were detected in the nuclear genome of *T. matsutake*, and several large ones were confirmed by PCR assays and read count comparisons (Supporting Figs. S2B & 3). How the evolution of NUMTs differs from mtDNA remains to be investigated in the future, and how NUMTs affect host nuclear genes is unknown. Transfer of mitochondrial DNA into the nucleus seems to be an

ongoing event. For fungi whose NUMTs were examined, NUMTs are commonly detected though at variable frequencies and different size distributions in different genomes (Hazkani-Covo *et al.*, 2010). For example, just few short NUMTs were detected in *Cordyceps militaris* (Zhang *et al.*, 2019b) and *Pestalotiopsis fici* (Zhang *et al.*, 2017a), while some large NUMTs were detected in *Tolypocladium inflatum* (Zhang *et al.*, 2017e) and *Glarea lozoyensis* (Zhang *et al.*, 2017d). Furthermore, in this study, the three examined *T. matsutake* samples also showed some differences. At least, those large NUMTs were not equally present in each sample (Supporting Fig. S2A). Therefore, *T. matsutake* may display intraspecific diversity on NUMT distribution, which is also another topic that deserves further investigation.

Evolutionary difference between nuclear and mitochondrial DNAs of *T. matsutake* was compared using whole mitogenomes and multilocus nuclear markers. Contrary to the greater variabilities of mtDNA than nuclear DNA frequently reported in animals (Lynch *et al.*, 2006), we found a higher mutation rates in nuclear DNA than mtDNA for *T. matsutake*. Actually, a growing body of literature suggests that, in many fungal species, there is less mitochondrial divergence than what is observed in the nuclear genome although the opposite is also true for certain fungal species (Sandor *et al.*, 2018). Phylogenetic topologies constructed based on mitochondrial and nuclear DNAs were significantly inconsistent for *T. matsutake* although mitochondrial and nuclear genetic distances were highly correlated (Supporting Tables S13 & S14). One possible explanation to the inconsistency between the two phylogenies could be their different inheritance patterns during hybridization in nature. Inheritance of

nuclear genomes follows the typical Mendelian principals of segregation and independent assortment. Although the mitochondrial inheritance of *T. matsutake* has not been fully investigated, it is likely to follow the dominant uniparental inheritance, being obviously different from the nuclear inheritance. Similar mitochondrial-nuclear evolutionary inconsistency was also reported in the entomopathogenic fungus *Cordyceps militaris* (Zhang *et al.*, 2019b).

This study also provides a basis to illuminate the origin and evolution of the fungus. Several independent studies have demonstrated that *T. matsutake* currently distributes in far eastern Asia (Japan, China, and Korea), northern Europe (Norway, Finland, and Sweden), and eastern North America (Chapela and Garbelotto, 2004; Sandor *et al.*, 2020). What factors have caused the sporadic distribution of this fungus? Where is the origin center of the fungus and how does it spread to other places? The putative uniparental inheritance of *T. matsutake* mitogenomes might provide an excellent opportunity to infer its evolutionary route. Our data support the origination of *T. matsutake* in Eurasia because of the higher genetic diversity of Eurasian samples than European samples. Specifically, the nine Eurasian samples belonged to nine different mito-types, and the nine European samples only showed two different mito-types (Fig. 2). The fungus might then disperse to North America via either the North Atlantic Bridge or the Bering Strait. Obviously, additional samples are needed to finally disclose the evolutionary route of the fungus. Our current study provides many useful markers for addressing these questions. Finally, it should also be noted that there was incongruency between sample geography and the population observed

based on either mitochondrial or nuclear molecular markers (Fig. 4). Samples failed to group according to their geography although the extent of genetic differentiation by geography remains to be examined. Factors driving genetic differentiation of the fungus need to be investigated in the future.

In conclusion, this study performed the first comparative mitogenomic study and greatly enriched the available genomic information of *T. matsutake* by performing high-throughput DNA sequencing of individuals originating from various regions. Comparison on different mitogenomes revealed low intraspecific variations in *T. matsutake* except for one *cox1* intron cox1P372 that showed presence/absence dynamics among different individuals. There was clear evidence of transfer of mitochondrial DNA into the nuclear genome. Mitochondrial DNA of *T. matsutake*, however, showed a lower mutation rate than nuclear DNA, and phylogenetic trees derived from mitochondrial and nuclear DNA sequences were incongruent. Together, this study revealed the dynamic genome evolution of the gourmet pine mushroom.

Experimental procedures

Sample information

Besides the three *T. matsutake* samples whose mitogenomes have been available in GenBank, 17 additional *T. matsutake* samples were used in this study (Table 1). These 20 samples originated from distant localities, including Canada (1 sample), China (5), Finland (6), Japan (3), South Korea (2), and Sweden (3) (Supporting Fig. S4). For 14 out of the 17 samples used in the current study, cultures were obtained from fruiting bodies and maintained on a modified MMN medium (Vaario *et al.*, 2010). Mycelia

were then collected and used for DNA extraction. For the remaining three samples (Tm1, Swe_2, and Tm_Korea), no mycelial cultures were available, and thus their fruiting bodies were directly used for DNA extraction. The 17 samples used in this study were all confirmed to be *T. matsutake* based on ITS phylogeny (Supporting Fig. S5).

Assembly and annotation of *T. matsutake* mitogenomes

Genomic DNAs extracted from cultures of the 14 samples were sequenced on an Illumina Xten platform in 2×150 bp reads after fragmenting by sonication to sizes of ~ 280 bp at Biomics (Beijing, China). Genomic DNAs extracted from fruiting bodies of other three samples were sequenced on an Illumina Hiseq platform in 2×90 bp reads after fragmenting by sonication to sizes of ~ 500 bp at BGI (Shenzhen, China). Reads with *N*s beyond 10% of read lengths and those containing low-quality bases (Q $\leqslant$ 20) exceeding 40% of read lengths were removed in pair, and remaining reads were considered clean data. Mitogenomes were *de novo* assembled from clean reads using NOVOPlasty (Dierckxsens *et al.*, 2017) and GetOrganelle (Jin *et al.*, 2020). Annotation of mitogenomes were performed as described previously (Zhang *et al.*, 2017e). Briefly, mitogenome sequences were first annotated by the online tool MFannot (https://megasun.bch.umontreal.ca/cgi-bin/dev_mfa/mfannotInterface.pl) based on the mold mitochondrial genetic code (i.e., genetic code 4), and then manual examination was performed to rectify any annotation errors that might exist. For intronic regions and intergenic regions longer than 300 bp, potential ORFs were searched using ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/) by setting

"ATG and alternative initiation codons" as start codons. Only ORFs with a minimal length of 300 nt were considered. Introns in *rnl* and protein-coding genes were determined and named according to proposals suggested by Johansen & Haugen (Johansen and Haugen, 2001) and Zhang & Zhang (Zhang and Zhang, 2019), respectively.

Repetitive elements in the *T. matsutake* mitogenome

In order to understand repetitive elements present in the *T. matutake* mitogenome, we performed four independent analyses using the mitogenome sequence of NF5, which was one of the largest mitogenomes and represented the dominant mitogenome-type among all *T. matsutake* samples. First, the NF5 mitogenome was searched against itself to detect potential intra-genome duplications using the NCBI BLAST web interface, optimizing for "more dissimilar sequences (discontiguous megablast)" under the default parameter settings. Hits with *E* values < 10^{-5} were included for further analyses. Second, we detected tandem repeats using Tandem Repeat Finder (Benson, 1999) under default settings. Third, we analyzed simple sequence repeats (SSR) using SSRIT (Temnykh *et al.*, 2001) by allowing the maximum motif-length group of decamer and the minimum repeat number of two. The SSRIT results were filtered to keep those with SSR length exceeding 10 bp, namely dimers with $\geqslant$ 5 repeats, trimers with $\geqslant$ 4 repeats, tetramers with $\geqslant$ 3 repeats, pentamers to decamers with $\geqslant$ 2 repeats. Finally, we used REPuter (Kurtz *et al.*, 2001) to identify forward (direct), reverse, complemented, and palindromic (revere complemented) repeats. The REPuter analysis was performed under default parameters except that

"Maximum Computed Repeats" was adjusted to 500, and repeats with *E* values < 10^{-5} were retained.

Comparison among different *T. matsutake* mitogenomes

All available complete mitogenomes of *T. matsutake* were adjusted to the same starting point (i.e., 35 bp upstream *rnl*) and then aligned using the online program MAFFT version 7 under default settings (Katoh *et al.*, 2019). According to annotations of the NF5 mitogenome, sequences of each gene were extracted from the multiple sequence alignment, and nucleotide variations at each gene were determined using DnaSP version 6.12.03 (Julio *et al.*, 2017). For intron-containing genes, nucleotide variations between intronic and exonic sequences were also compared.

A phylogenetic tree consisting of all *T. matsutake* samples was constructed based on whole mitogenome sequences by the Maximum Parsimony (MP) method as implemented in PAUP version 4.0b10 (Swofford, 2003). Alignment gaps were defined as the fifth base, and the tree was mid-point rooted. To further provide a framework for understanding the evolution of the *cox1* intron cox1P372, an MrBayes tree was constructed using concatenated exon sequences of 14 core protein-coding genes (*nad1-6*, *4L*; *cob*; *cox1-3*, and *atp6*, *8*, *9*) and two rRNA genes (*rnl* and *rns*) from 11 representative mitogenomes of *T. matsutake* as well as two outgroups, *Tricholoma flavovirens* (MN873037) and *Tricholoma saponaceum* (MN873038), using the same protocol as that described previously(Zhang *et al.*, 2014). The generated MrBayes tree was used to infer the ancestral intron state by defining two states, A for possession of cox1P372 and B for deletion of cox1P372, using RASP as described previously

(Wang *et al.*, 2018).

Detection of mitochondrial heteroplasmy

Heteroplasmy refers to the presence of different mitochondrial genomes in the same individual. While each matsutake strain has two copies of each nuclear gene per cell, the mitochondrial genome copy numbers are usually much higher and can be different among copies. Here, we examined potential evidence for heteroplasmy for samples with completely assembled mitogenomes. Heteroplasmy detection was implemented by NOVOPlasty (Dierckxsens *et al.*, 2019). Different from the normal assembly mode of the program, in the heteroplasmy mode, the previously assembled complete mitogenome of each sample was used simultaneously as reference and as seed inputs, and the minimum minor allele frequency (MAF) was set to 0.03, where allelic variants with lower frequencies than 0.03 were ignored. To confirm putative heteroplasmic sites, PCR primers were designed at adjacent regions flanking target sites, and amplicons were checked by Sanger sequencing.

Detection of nuclear mitochondrial DNA segments (NUMTs)

To detect the potential transfer of mitochondrial genes to the nuclear genomes in matsutake mushrooms, we first assembled the genomes for three representative samples, NF5, F2 and L1, using IDBA-UD (Peng *et al.*, 2012). Local BLAST analyses of each mitogenome were performed against assembled contigs of corresponding samples, and those contigs displaying identities > 99.5% with the mitogenome at coverages > 99.5% were regarded as mitochondrial ones. The remaining contigs were deemed as those in the nuclear genome. The mitogenome of

each sample was then BLAST searched against their corresponding nuclear assemblies using two approaches, BLAST (Altschul *et al.*, 1990) and LAST (Kiełbasa *et al.*, 2011). The results were filtered to keep those with *E* values < 10^{-5} and identities > 80%. The authenticity of nuclear contigs having an alignment length of > 500 nt with mitogenomes were verified by PCR surveys (Supplementary Table S15). For each PCR survey, one PCR primer was designed within the matching region (i.e., within NUMTs), and the other PCR primer was designed at regions upstream or downstream NUMTs of corresponding nuclear contig. The latter primers were ensured to have no homology to mitogenomes (known by online BLAST) in order to increase amplification specificity. Sanger sequencing were performed for amplicons of expected sizes.

Comparison on divergence between mitochondrial and nuclear DNAs

In order to compare the divergence between nuclear and mitochondrial DNAs, we investigated GenBank records of *T. matsutake* and determined 12 nuclear markers, including nine protein-encoding genes of dioxygenase (*dox1*, GenBank accession no. LC582843), endoglucanase (*Egl5A*, LC424191), O-methyltransferase (*omt1*, LC417117), glycoside hydrolase (*Amy1*, LC367230), actin (*act*, LC011390), laccase (*lac*, DI377006), phenylalanine ammonia-lyase (*PAL2*, AB901372), glyceraldehyde 3 phosphate dehydrogenase (*gpd*, JQ696991), and translation elongation factor 1-alpha (*tef*, AB699731), and three fragments (ITS, *nrSSU*, and *nrLSU*; U62538, U62964) of nuclear rDNA. Sequences of these nuclear markers for each sample were extracted from their whole genome assemblies, which were obtained using IDBA-UD, through

local BLAST searches. Indices of nucleotide diversity and haplotype diversity at concatenated nuclear fragments were estimated using DnaSP version 6.12.03 (Julio *et al.*, 2017) and then compared to indices at mitogenomes. Nucleotide diversity values indicate the number of nucleotide differences for each site, on average, that exist between two sequences selected at random. Haplotype diversities provided a probability that two randomly selected sequences are different.

A phylogenetic tree was constructed based on concatenated nuclear markers using the MP method described above. In order to directly observe and compare topologies constructed based on mitogenomes and nuclear gene fragments, a tanglegram was constructed using TREEMAP 3b124328(Charleston, 2011). To test whether mitogenomes and nuclear genomes evolve congruently or independently and whether their topological difference was statistically significant, we performed three statistical tests: the parsimony-based Kishino-Hasegawa (KH) test, the likelihood-based Shimodaira-Hasegawa (SH) test, and the partition homogeneity test (PHT) in PAUP as described previously(Zhang *et al.*, 2015). In addition, Mantel test was performed to understand the correlation between different distance matrices with the program ZT (Bonnet and de Peer, 2002).

Data Accessibility

The complete *T. matsutake* mitogenome sequences newly generated in this study were submitted to GenBank under accession numbers MZ042935, MZ066593–MZ066607. NUMT sequences were submitted to GenBank under accession numbers OK137539–OK137544. Clean reads were deposited in the NCBI Sequence Read Archive (SRA)

database under the accession number PRJNA726361.

Author Contributions

Y.J.Z., J.P.X., and L.M.V. designed research; Y.J.Z., and S.Z. performed research; Y.J.Z., X.B., L.Y.R., H.H.S. and H.P.T analyzed data; S.Z., Y.J.Z., J.P.X. and L.M.V. wrote the paper.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgements

This study was funded by the National Natural Science Foundation of China (31872162) and Hundred Talents Program of Shanxi Province. We thank Hong-Yue Zhang for helping cultivate *T. matsutake* strains and collect mycelia, and the High Performance Simulation Platform of Shanxi University for providing computing resource.

References

Adams, K., and Palmer, J.D. (2003) Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol* **29**: 380-395.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**: 573-580.

Bonnet, E., and de Peer, Y.V. (2002) zt: A Sofware Tool for Simple and Partial Mantel Tests. *J Stat Softw* **7**: 1-12.

Brandrud, T.-E. (2020) *Tricholoma matsutake* (amended version of 2020 assessment). *The IUCN Red List of Threatened Species* **2020**: e.T76267712A177054645.

Chapela, I.H., and Garbelotto, M. (2004) Phylogeography and evolution in matsutake and close allies inferred by analyses of ITS sequences and AFLPs. *Mycologia* **96**: 730-741.

Charleston, M.A. (2011). TreeMap 3b. URL https://sites.google.com/site/cophylogeny

Chatre, L., and Ricchetti, M. (2014) Are mitochondria the Achilles' heel of the Kingdom Fungi? *Curr Opin Microbiol* **20**: 49-54.

Dierckxsens, N., Mardulyn, P., and Smits, G. (2017) NOVOPlasty: de novo assembly of organelle

genomes from whole genome data. *Nucleic Acids Res* **45**: e18.
Dierckxsens, N., Mardulyn, P., and Smits, G. (2019) Unraveling heteroplasmy patterns with NOVOPlasty. *NAR Genom Bioinf* **2**: 1-10.
Fan, W.-W., Zhang, S., and Zhang, Y.-J. (2019) The complete mitochondrial genome of the Chan-hua fungus *Isaria cicadae*: a tale of intron evolution in Cordycipitaceae. *Environ Microbiol* **21**: 864-879.
Hazkani-Covo, E., Zeller, R.M., and Martin, W. (2010) Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet* **6**: e1000834.
Huang, W., Feng, H., Tu, W., Xiong, C., Jin, X., Li, P. *et al.* (2021) Comparative mitogenomic analysis reveals dynamics of intron within and between *Tricholoma* species and phylogeny of Basidiomycota. *Front Genet* **12**: 534871.
Isaac, R.S., McShane, E., and Churchman, L.S. (2018) The multiple levels of mitonuclear coregulation. *Annu Rev Genet* **52**: 511-533.
Jin, J.-J., Yu, W.-B., Yang, J.-B., Song, Y., Yi, T.-S., and Li, D.-Z. (2020) GetOrganelle: a simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *Genome Biol* **21**: 241.
Johansen, S., and Haugen, P. (2001) A new nomenclature of group I introns in ribosomal DNA. *RNA* **7**: 935-936.
Julio, R., Albert, F.M., Carlos, S.-D.J., Sara, G.R., Pablo, L., Ramos-Onsins, S.E., and Alejandro, S.G. (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol* **34**: 3299–3302.
Katoh, K., Rozewicki, J., and Yamada, K.D. (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* **20**: 1160-1166.
Kiełbasa, S.M., Wan, R., Sato, K., Horton, P., and Frith, M.C. (2011) Adaptive seeds tame genomic sequence comparison. *Genome Res* **21**: 487-493.
Kurokochi, H. (2021) Genetic structure of *Tricholoma matsutake* in Japan: conservation of genetic resources of domestic "matsutake" mushrooms. *J Forest Res* **26**: 62-67.
Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* **29**: 4633-4642.
Lynch, M., Koskella, B., and Schaack, S. (2006) Mutation pressure and the evolution of organelle genomic architecture. *Science* **311**: 1727-1730.
Min, B., Yoon, H., Park, J., Oh, Y.-L., Kong, W.-S., Kim, J.-G., and Choi, I.-G. (2020) Unusual genome expansion and transcription suppression in ectomycorrhizal *Tricholoma matsutake* by insertions of transposable elements. *PloS one* **15**: e0227923.
Nie, Y., Wang, L., Cai, Y., Tao, W., Zhang, Y.-J., and Huang, B. (2019) Mitochondrial genome of the entomophthoroid fungus *Conidiobolus heterosporus* provides insights into evolution of basal fungi. *Appl Microbiol Biotechnol* **103**: 1379-1391.
Peng, Y., Leung, H.C., Yiu, S.M., and Chin, F.Y. (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420-1428.
Sandor, S., Zhang, Y.-J., and Xu, J. (2018) Fungal mitochondrial genomes and genetic polymorphisms. *Appl Microbiol Biotechnol* **102**: 9433-9448.
Sandor, S.R., Wang, H., Vaario, L., Trudell, S.A., and Xu, J. (2020) Mitochondrial multilocus DNA sequence analyses reveal limited genetic variability within and consistent differences between species of the global matsutake species complex. *Acta Edulis Fungi* **27**: 1-19.
Swofford, D.L. (2003). PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4.0 b10. URL http://paup.phylosolutions.com/

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., and McCouch, S. (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res* **11**: 1441-1452.

Vaario, L.-M., Yang, X., and Yamada, A. (2017) Biogeography of the Japanese gourmet fungus, *Tricholoma matsutake*: a review of the distribution and functional ecology of matsutake. *Ecol Stud* **230**: 319-344.

Vaario, L.M., Pennanen, T., Sarjala, T., Savonen, E.-M., and Heinonsalo, J. (2010) Ectomycorrhization of *Tricholoma matsutake* and two major conifers in Finland—an assessment of in vitro mycorrhiza formation. *Mycorrhiza* **20**: 511-518.

Wang, Yun, Yu, Fuqiang, Zhang, Chunxiang *et al.* (2017a) *Tricholoma matsutake*: an edible mycorrhizal mushroom of high socioeconomic relevance in China. *Rev Mex Micol* **46**: 55-61.

Wang, L., Zhang, S., Li, J.-H., and Zhang, Y.-J. (2018) Mitochondrial genome, comparative analysis and evolutionary insights into the entomopathogenic fungus *Hirsutella thompsonii*. *Environ Microbiol* **20**: 3393-3405.

Wang, P., Sha, T., Zhang, Y., Cao, Y., Mi, F., Liu, C. *et al.* (2017b) Frequent heteroplasmy and recombination in the mitochondrial genomes of the basidiomycete mushroom *Thelephora ganbajun*. *Sci Rep* **7**: 1626.

Winkler, D. (2008) The mushrooming fungi market in Tibet -- exemplified by *Cordyceps sinensis* and *Tricholoma matsutake*. *J Int Assoc Tibet Stud*: 1-47.

Woodson, J.D., and Chory, J. (2008) Coordination of gene expression between organellar and nuclear genomes. *Nat Rev Genet* **9**: 383-395.

Xu, J., and Li, H. (2015) Current perspectives on mitochondrial inheritance in fungi. *Cell Health Cytoskel* **7**: 143-154.

Xu, J., and Wang, P. (2015) Mitochondrial inheritance in basidiomycete fungi. *Fungal Biol Rev* **29**: 209-219.

Xu, J., Cadorin, M., Liang, Y.-J., and Yang, Z.L. (2010) DNA-based geographic typing of the gourmet mushroom *Tricholoma matsutake* traded in China. *Mycoscience* **51**: 248-251.

Xu, J., Sha, T.A.O., Li, Y.-C., Zhao, Z.-W., and Yang, Z.L. (2008) Recombination and genetic differentiation among natural populations of the ectomycorrhizal mushroom *Tricholoma matsutake* from southwestern China. *Mol Ecol* **17**: 1238-1247.

Yamanaka, T., Yamada, A., and Furukawa, H. (2020) Advances in the cultivation of the highly-prized ectomycorrhizal mushroom *Tricholoma matsutake*. *Mycoscience* **61**: 49-57.

Yoon, H., Kong, W.-S., Kim, Y.J., and Kim, J.-G. (2016) Complete mitochondrial genome of the ectomycorrhizal fungus *Tricholoma matsutake*. *Mitochondr DNA A* **27**: 3855-3857.

Yue, L.-l., Wang, X., and Zhang, L.-Z. (2017) The research progress of *Tricholoma matsutake*. *Spec Wild Econ Anim Plant Res* **39**: 72-76.

Zeng, D.-F., and Chen, B. (2015) Genetic variability and bottleneck detection of four *Tricholoma matsutake* populations from northeastern and southwestern China. *Environ Microbiol* **17**: 2870-2881.

Zhang, S., and Zhang, Y.-J. (2019) Proposal of a new nomenclature for introns in protein-coding genes in fungal mitogenomes. *IMA Fungus* **10**: 15.

Zhang, S., Zhang, Y., and Li, Z. (2019a) Complete mitogenome of the entomopathogenic fungus *Sporothrix insectorum* RCEF 264 and comparative mitogenomics in Ophiostomatales. *Appl Microbiol Biotechnol* **103**: 5797-5809.

Zhang, S., Cui, N., Zhao, Y., and Zhang, Y. (2019b) Comparison of evolutionary relationships between

mitochondrial and nuclear DNAs of *Cordyceps militaris*. *Acta Microbiol Sin* **59**: 2346-2356.
Zhang, S., Wang, X.-N., Zhang, X.-L., Liu, X.-Z., and Zhang, Y.-J. (2017a) Complete mitochondrial genome of the endophytic fungus *Pestalotiopsis fici*: features and evolution. *Appl Microbiol Biotechnol* **101**: 1593–1604.
Zhang, S., Hao, A.J., Zhao, Y.X., Zhang, X.Y., and Zhang, Y.-J. (2017b) Comparative mitochondrial genomics toward exploring molecular markers in the medicinal fungus *Cordyceps militaris*. *Sci Rep* **7**: 40219.
Zhang, Y.-J., Zhang, H.-Y., Liu, X.-Z., and Zhang, S. (2017c) Mitochondrial genome of the nematode endoparasitic fungus *Hirsutella vermicola* reveals a high level of synteny in the family Ophiocordycipitaceae. *Appl Microbiol Biotechnol* **101**: 3295-3304.
Zhang, Y.-J., Zhao, Y.-X., Zhang, S., Chen, L., and Liu, X.-Z. (2017d) Reanalysis of the mitochondrial genome of the pneumocandin-producing fungus *Glarea lozoyensis*. *Acta Microbiol Sin* **57**: 724-736.
Zhang, Y.-J., Yang, X.-Q., Zhang, S., Humber, R.A., and Xu, J. (2017e) Genomic analyses reveal low mitochondrial and high nuclear diversity in the cyclosporin-producing fungus *Tolypocladium inflatum*. *Appl Microbiol Biotechnol* **101**: 8517-8531.
Zhang, Y.-J., Zhang, S., Zhang, G., Liu, X., Wang, C., and Xu, J. (2015) Comparison of mitochondrial genomes provides insights into intron dynamics and evolution in the caterpillar fungus *Cordyceps militaris*. *Fungal Genet Biol* **77**: 95-107.
Zhang, Y.-J., Zhang, S., Li, Y., Ma, S., Wang, C., Xiang, M. *et al.* (2014) Phylogeography and evolution of a fungal-insect association on the Tibetan Plateau. *Mol Ecol* **23**: 5337-5355.

Table 1 Origin and mitogenome information of *T. matsutake* samples

Sample	DNA source	Year ‡	Origin	Dominant host tree ‡	Mitogenome size (bp)	Accession no.	Note
JT5	Culture	N/A	Kyoto, Japan	*Pinus densiflora*	76071	MZ066600	
JT89	Culture	N/A	Nagano, Japan	*Pinus densiflora*	76067	MZ066601	
LC385608 †	Culture	1993	Ibaraki, Japan	N/A	76089	LC385608	
F2	Culture	2002	Fuyu, Jilin, China	*Pinus densiflora*	76073	MZ066602	Northeast China
W2	Culture	2002	Wangqing, Jilin, China	*Pinus densiflora*	76074	MZ066603	Northeast China
L1	Culture	2003	Lufeng, Yunnan, China	*Castanopsis orthacantha*	74750	MZ066605	Southwest China
TM-NH	Culture	2016	Nanhua, Yunnan, China	N/A	74767	MZ066607	Southwest China
MN873034 †	Fruiting body	N/A	Yajiang, Sichuan, China	N/A	74865	MN873034	Southwest China
Tm_Korea	Fruiting body	N/A	South Korea	N/A	Incomplete	--	
JX985789 †	Fruiting body	N/A	Gachang, South Korea	N/A	76037	JX985789	
EF	Culture	2007	Kontiolahti, Finland	*Pinus sylvestris*	74750	MZ066604	
NF5	Culture	2010	Rovaniemen, Finland	*Pinus sylvestris*	76034	MZ066595	
NF6	Culture	2010	Rovaniemen, Finland	*Pinus sylvestris*	76034	MZ066596	
N18	Culture	2008	Espoo, Finland	*Pinus sylvestris , Picea abies*	76034	MZ066594	
T8	Culture	2012	Espoo, Finland	*Pinus sylvestris*	76034	MZ066599	
C4	Culture	2013	Espoo, Finland	*Pinus sylvestris*	76034	MZ066593	
SU3	Culture	2010	Umeå, Sweden	*Pinus sylvestris*	76034	MZ066597	
SW4	Culture	2014	Klaxi, Sweden	*Pinus sylvestris*	76034	MZ066598	
Swe_2	Fruiting body	N/A	Vagnharad, Sweden	N/A	76034	MZ042935	
Tm1	Fruiting body	N/A	Quebec, Canada	N/A	74802	MZ066606	

† The three samples downloaded from GenBank are represented by their accession nos. ‡ N/A, not available.

Table 2 Features of introns and intronic ORFs characterized in the *T. matsutake* NF5 mitogenome

Intron/intronic ORF	Start	End	Length (bp)	Intron type	Standard name	Start codon †	Stop codon	Encoded protein
rnl-i1	4885	5922	1038	IA	mL2585			
orf235	4910	5617	708			ATG	TAA	LAGLIDADG
cox1-i1	24958	26258	1301	IB (3', partial)	cox1P281			
orf290	24959	25831	873			ND	TAA	LAGLIDADG
cox1-i2	26350	27639	1290	IB (complete)	cox1P372			
orf146	26856	27296	441			ATG	TAA	LAGLIDADG
cox1-i3	27654	28893	1240	IB (complete)	cox1P386			
orf360	27655	28737	1083			ND	TAA	LAGLIDADG
cox1-i4	29228	30321	1094	IB (complete)	cox1P720			
orf343	29228	30259	1032			ND	TAG	LAGLIDADG
cox1-i5	30502	31672	1171	IB (extra insertion)	cox1P900			
orf338	30502	31518	1017			ND	TAA	LAGLIDADG
nad1-i1	45967	47472	1506	IB (complete)	nad1P276			N/A ‡
cob-i1	52238	53526	1289	IB (complete)	cobP823			N/A ‡
cox2-i1	55328	56489	1162	IB (complete)	cox2P318			N/A ‡
cox2-i2	56529	59129	2601	ID	cox2P357			N/A ‡
orf378	56529	57665	1137			ND	TAA	LAGLIDADG
nad5-i1	62481	63529	1049	IB (complete)	nad5P417			

† ND, start codon not determined in corresponding intron sequences. In most cases, start codons of these ORFs could be identified in the upstream exon sequences.

‡ N/A, no ORF detected.

Table 3 Nucleotide variations at different regions among 19 *T. matsutake* mitogenomes.

Item in comparison	Length (nt)	Pi	S	Indel	Total (nt)	SNP%	Overall%
Whole sequence (including cox1P372)	76371	52	68	1800	1920	0.16	2.51
Whole sequence after excluding cox1P372	75081	52	68	510	630	0.16	0.84
Intergenic regions	33406	35	49	355	439	0.25	1.31
Genic regions (excluding cox1P372)	41675	17	19	155	191	0.09	0.46
Exonic regions (only for intron-containing genes)	11793	5	5	6	16	0.08	0.14
Intronic regions (including cox1P372)	14758	9	9	19	37	0.12	0.25
rnl exons	5292	2	4	6	12	0.11	0.23
rnl-i1 (mL2585)	1038	0	0	0	0	0.00	0.00
cox1 exons	1584	0	0	0	0	0.00	0.00
cox1-i1 (cox1P281)	1302	2	1	1	4	0.23	0.31
cox1-i2 (cox1P372)	1290	0	0	0	0	0.00	0.00
cox1-i3 (cox1P386)	1240	0	0	0	0	0.00	0.00
cox1-i4 (cox1P720)	1094	3	0	0	3	0.27	0.27
cox1-i5 (cox1P900)	1171	1	1	0	2	0.17	0.17
nad1 exons	999	0	0	0	0	0.00	0.00
nad1-i1 (nad1P276)	1507	2	1	1	4	0.20	0.27
cob exons	1161	1	1	0	2	0.17	0.17
cob-i1 (cobP823)	1294	1	0	5	6	0.08	0.46
cox2 exons	762	1	0	0	1	0.13	0.13
cox2-i1 (cox2P318)	1169	0	3	7	10	0.26	0.86
cox2-i2 (cox2P357)	2602	0	2	3	5	0.08	0.19
nad5 exons	1995	1	0	0	1	0.05	0.05
nad5-i1 (nad5P417)	1051	0	1	2	3	0.10	0.29
Sum. PCG exons (only for intron-containing genes)	6501	3	1	0	4	0.06	0.06
Sum. PCG introns	13720	9	9	19	37	0.13	0.27

Table 4 Detection of putative heteroplasmy in mitogenomes of different *T. matsutake* samples †

Sample	Position	Base(Ref/Alt)	AF	DP	FR	LCR
C4	66178	AATA/A	0.034	828	R	
N18	17424	TAT/T	0.035	1448	R	
NF5	17424	TAT/T	0.036	386	R	
NF6	17426	T/TAT	0.031	320	R	
SU3	9421	G/T	0.067	89	FR	
SW4	17426	T/TAT	0.037	326	R	
Swe_2	1141	T/G	0.041	394	R	
	4846	T/G	0.032	286	R	
	16201	A/G	0.042	381	R	SNR
	16212	G/A	0.035	572	R	
	16220	C/T	0.043	605	R	
T8	**682**	**C/T**	**0.058**	**39**	**F**	
	830	**AA/A**	**0.031**	**96**	**R**	
JT5	No detection					
JT89	**17426**	**TAT/T**	**0.041**	**290**	**R**	
	17429	**T/A**	**0.035**	**196**	**F**	**SNR**
	66209	AATA/A	0.044	180	R	
F2	17425	TAT/T	0.036	442	R	
	58703	G/T	0.033	628	R	SNR
	68227	A/T	0.050	473	R	SNR
W2	17425	TAT/T	0.036	442	R	
	51118	TAAT/T	0.031	258	R	
	68228	A/T	0.032	253	R	SNR
EF	4608	G/A	0.038	104	FR	
	9988	T/G	0.070	129	F	
	17421	TAT/T	0.071	84	R	
	49329	**G/T**	**0.041**	**118**	**F**	
	49824	**TAAT/T**	**0.050**	**60**	**R**	
	66917	A/-		71	R	SNR
	71199	C/T	0.080	136	F	
L1	17421	TAT/T	0.032	438	R	
	17424	T/A	0.060	279	F	SNR
Tm1	20494	G/T	0.063	328	R	SNR
	20500	C/T	0.196	236	R	SNR
	20504	C/T	0.146	180	R	SNR
	35436	G/A	0.074	515	FR	SNR(A)
TM_NH	No detection					

† AF, allele frequency; DP, depth of coverage; FR, detected on the forward (F) and/or reverse (R) strand; LCR, low complexity region; SNR, single-nucleotide repeat. Sites that were examined by PCR were shown in bold.

Table 5 Nucleotide variations on different nuclear markers among 16 *T. matsutake* samples †

Gene	Length (nt)	Pi	S	Indel	Total (nt)	SNP %	Overall %
dox1	5108	13	10	2	25	0.45	0.49
Egl5A	2023	2	6	0	8	0.40	0.40
Omt1	2059	7	6	0	13	0.63	0.63
Amy1	2226	27	7	9	43	1.53	1.93
act	676	0	0	0	0	0.00	0.00
lac	2093	8	9	0	17	0.81	0.81
PAL2	2983	26	7	7	40	1.11	1.34
gpd	680	2	5	0	7	1.03	1.03
tef	465	1	0	0	1	0.22	0.22
ITS	601	1	3	10	14	0.67	2.33
nrLSU	3410	2	0	0	2	0.06	0.06
nrSSU	1807	0	0	0	0	0.00	0.00
Overall	24131	89	53	28	170	0.59	0.70
Overall w/o *act1* & *nrSSU*	21648	89	53	28	170	0.66	0.79

† In this table, only the 16 samples with complete mitogenomes from the current study were used.

Figure legends

Fig. 1 Circular map of the *T. matsutake* NF5 mitogenome (A) and composition of different elements (B). Different elements were shown in different colours. Genes transcribed in the forward strand were shown in the outward side of the ring, and genes transcribed in the reverse strand were shown in the inward side of the ring. Intron-containing genes were followed by an asterisk after gene names. Introns in panel B included introns themselves and intronic ORFs.

Fig. 2 The MP phylogenetic tree of 19 *T. matsutake* samples based on whole mitogenomes (A) and inference of ancestral distribution pattern of cox1P372 (B). In panel A, the 19 mitogenomes clustered into two clades (Pop1 and Pop2). Mito-types, geographical origins and the cox1P372 states (yes for presence and no for absence) were marked for each sample. In panel B, letters A and B in terminal tips and internal nodes indicated presence and absence of cox1P372, respectively.

Fig. 3 Comparison among different *T. matsutake* mitogenomes. Rings from inside to outside represented mito-types from A to K, consistent to those defined in Fig. 2A. Mito-types A-G mainly differed from mito-types H-K with regard to presence/absence of cox1P372.

Fig. 4 Comparison of the tanglegram generated by nuclear and mitochondrial markers. Mito-types (right side) were indicated using different letters, and nuclear haplotypes (left side) were indicated using circled numbers. Bootstrap support values were indicated at internal nodes using underlined numbers. Sample locations were

indicated above connecting lines. Fin, Finland; Swe, Sweden; Jpn, Japan; Can, Canada; NEC, Northeast China; SWC, Southwest China.

Figure 1

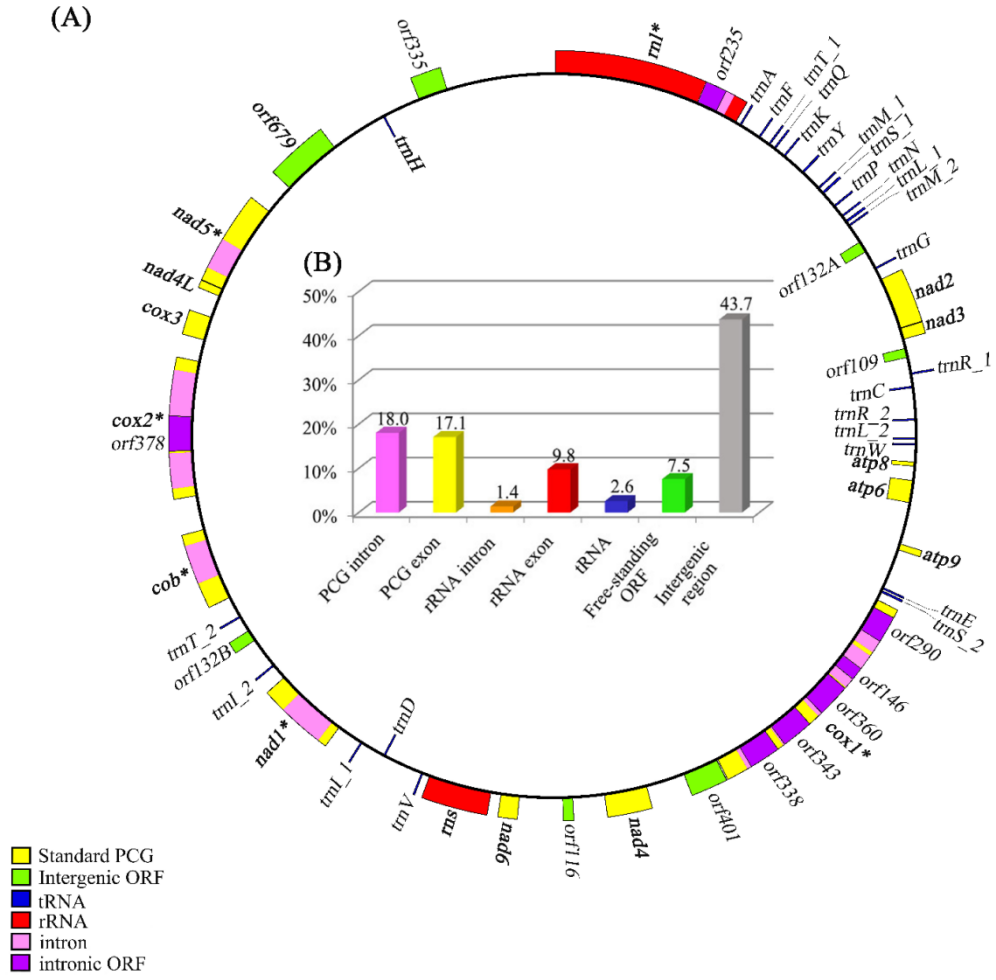

Figure 2

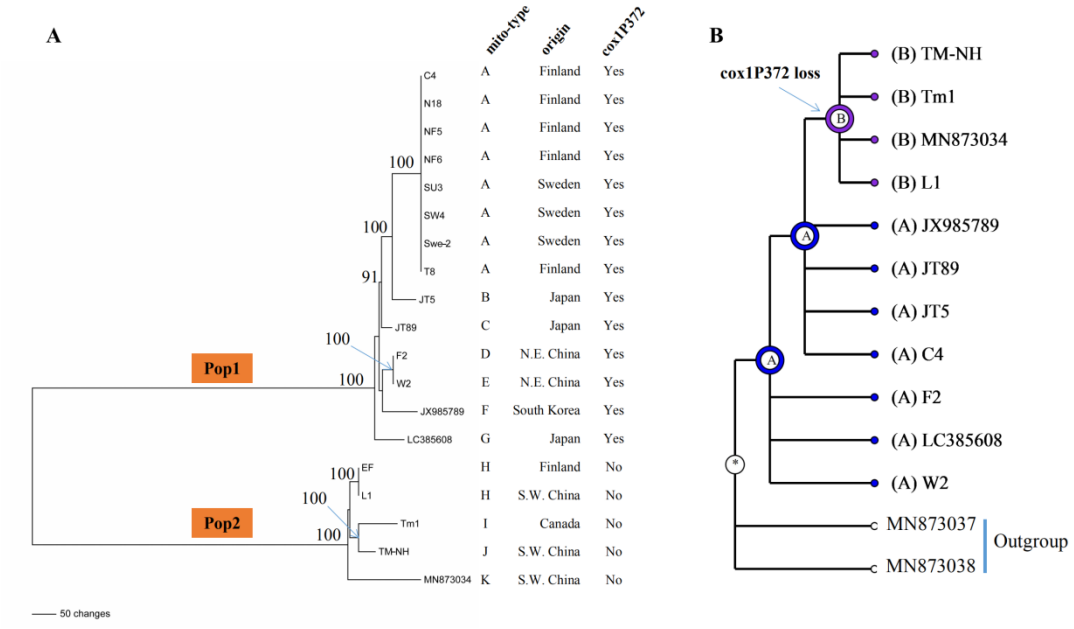

Figure 3

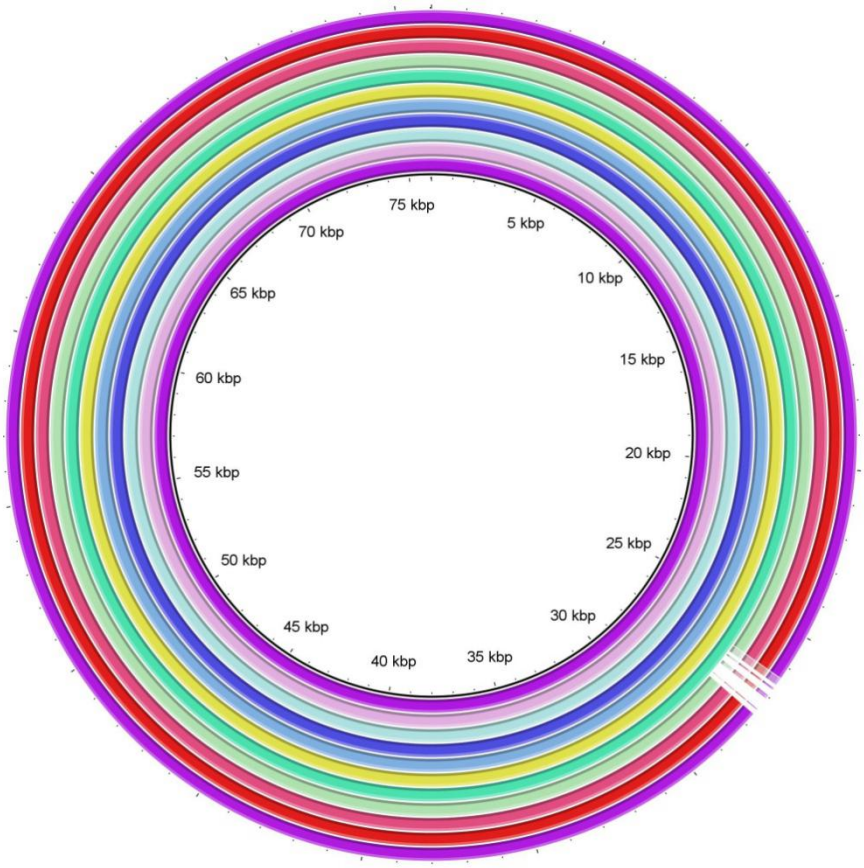

Figure 4