

Enrichment of G4DNA and a Large Inverted Repeat Coincide in the Mitochondrial Genomes of *Termitomyces*

Mathijs Nieuwenhuis^{1,*}, Lennart J.J. van de Peppel¹, Freek T. Bakker², Bas J. Zwaan¹, and Duur K. Aanen¹

¹Laboratory of Genetics, Wageningen University & Research, The Netherlands

²Biosystematics Group, Wageningen University & Research, The Netherlands

*Corresponding author: E-mail: mathijs.nieuwenhuis@wur.nl.

Accepted: June 11, 2019

Data deposition: Sequence data for this project has been deposited at NCBI GenBank under the accessions MG783568, MH725791-MH725800, and MH743217.

Abstract

Mitochondria retain their own genome, a hallmark of their bacterial ancestry. Mitochondrial genomes (mtDNA) are highly diverse in size, shape, and structure, despite their conserved function across most eukaryotes. Exploring extreme cases of mtDNA architecture can yield important information on fundamental aspects of genome biology. We discovered that the mitochondrial genomes of a basidiomycete fungus (*Termitomyces* spp.) contain an inverted repeat (IR), a duplicated region half the size of the complete genome. In addition, we found an abundance of sequences capable of forming G-quadruplexes (G4DNA); structures that can disrupt the double helical formation of DNA. G4DNA is implicated in replication fork stalling, double-stranded breaks, altered gene expression, recombination, and other effects. To determine whether this occurrence of IR and G4DNA was correlated within the genus *Termitomyces*, we reconstructed the mitochondrial genomes of 11 additional species including representatives of several closely related genera. We show that the mtDNA of all sampled species of *Termitomyces* and its sister group, represented by the species *Tephrocybe rancida* and *Blastosporella zonata*, are characterized by a large IR and enrichment of G4DNA. To determine whether high mitochondrial G4DNA content is common in fungi, we conducted the first broad survey of G4DNA content in fungal mtDNA, revealing it to be a highly variable trait. The results of this study provide important direction for future research on the function and evolution of G4DNA and organellar IRs.

Key words: fungi, mtDNA, G-quadruplex, inverted repeat, Lyophyllaceae.

Introduction

Mitochondria are a vital component of virtually all eukaryotic life, generating energy in the form of ATP and regulating the respiratory metabolism. Mitochondria maintain their own genome, a legacy of their endosymbiotic origin. However, despite their common ancestry and conserved function, mitochondrial genomes (mtDNA) of major lineages such as plants, fungi, and metazoans differ considerably not only in size, content, and structure but also in mutability, inheritance, and replication (Birky 2001; Burger et al. 2003; Lynch et al. 2006; Wilson and Xu 2012; Smith and Keeling 2015). Fungi by themselves contain a huge variety of mtDNA shapes and sizes, and are an opportune group for studying the causes and effects of mtDNA diversity.

Like the vast majority of eukaryotes, in most fungi mtDNA transmission during sexual reproduction is uniparental.

Species of the genus *Termitomyces* (Lyophyllaceae, Basidiomycota), a clade of obligate mutualistic symbionts of termites, are an exception. During mating in *Termitomyces*, nuclei of both parents enter a shared cytoplasm from which offspring cells grow, rather than one parent adopting a nucleus from the other as is typical for most basidiomycetes (Nobre et al. 2014). This implies that fertilized cells initially are heteroplasmic (containing the mitochondria of more than one parent), increasing the potential for recombination of mtDNA. This is similar to what has been demonstrated in *Agaricus bisporus* (Xu et al. 2013), another fungus that reproduces in this fashion. These deviations from “canonical,” uniparentally inherited, nonrecombining mtDNA render these fungi of high interest for studies on mtDNA evolution, as they enable testing hypotheses on possible influence of recombination on mtDNA selection and evolution. However, to

© The Author(s) 2019. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

date, no mtDNA sequence for *Termitomyces* or any closely related fungus has been published. In this article, we reveal the mtDNA sequences of seven species of *Termitomyces*, showing they have a mitochondrial inverted repeat (IR) that generally occupies half of the genome, duplicating several key genes and increasing the mtDNA size significantly. In addition, we found an abundance of G-quadruplex (G4DNA) motifs, prompting the question whether these two structural features are functionally correlated. We also sequenced the mtDNA of five other Lyophyllaceae: *Blastosporella zonata*, *Tephrocybe rancida*, *Myochromella boudieri*, *Asterophora parasitica*, and *Tricholomella constricta*, to estimate the origin of the IR and the G4DNA increase.

G-quadruplex DNA, or G4DNA, is a naturally occurring conformation of DNA that folds in a four-stranded conformation rather than the common double helix. Although originally considered as an *in vitro* curiosity, G4DNA is now known to exist in living cells of ciliates, humans, yeasts, nematodes, and others (Schaffitzel et al. 2001; Johnson et al. 2008; Verma et al. 2009; Koole et al. 2014). The formation of G4DNA requires between one and four strands of DNA containing tandemly repeated guanine motifs. The shape and stability of G4DNA depends on the number of comprising strands and their nucleotide sequence. G4DNA complexes can be stabilized or destabilized by specific binding proteins, particularly helicases (Mendoza et al. 2016).

G4DNA can have many effects depending on the location at which they occur and the shape they form. Effects associated with G4DNA include: replication fork stalling, altered gene expression, recombination, double-stranded break (DSB) formation, and genome instability (Maizels 2006; Rawal et al. 2006; Verma et al. 2009; Bharti et al. 2014; Huang et al. 2015; Cui et al. 2016; Rawal et al. 2006). In addition, the telomeres of many eukaryotes comprise G4DNA forming motifs. G4DNA formation at the ends of telomeres regulates telomerase activity (Zaug et al. 2005). G4DNA is also enriched around replication origins in vertebrates, and is considered to play a role in genome replication (Paeschke et al. 2011; Valton et al. 2014). G4DNA forms more frequently in cancer cells (Biffi et al. 2014), and G4DNA motifs are prevalent in promoters of oncogenes (Eddy and Maizels 2006). Given these properties, most research on G4DNA focuses on its potential role in cancer development (Hänsel-Hertsch et al. 2017).

Recent studies have employed predictive algorithms to map potential G4DNA in sequenced genomes based on sequence patterns (Eddy and Maizels 2006; Rawal et al. 2006; Du et al. 2008; Yadav et al. 2008; Capra et al. 2010; Garg et al. 2016). This revealed a plethora of G4DNA motifs in human, eumetazoan, plant, and yeast DNA. In human mitochondria, G4DNA motifs were found in close proximity to regions of instability related to mitochondrial dysfunction and diseases (Bharti et al. 2014). Capra et al. (2010) analyzed the G4DNA motif content of the nuclear and mitochondrial

genome of *Saccharomyces cerevisiae*, and found an ~10-fold increase of G4DNA motif content in mtDNA compared with nuclear DNA. However, a broad, comparative survey of G4DNA content of fungal nuclear and mtDNA is still lacking.

Large inverted duplications are a common, ancestral feature in chloroplast genomes (Palmer 1985; Turmel et al. 1999). They generally drive a process of continuous homologous recombination (HR) which maintains sequence equality between the two repeat copies, and inverts the orientation of the nonduplicated regions (Palmer 1983; Aldrich et al. 1985). IRs are comparatively rare in mtDNA, being a derived character opposed to the chloroplast IR, although mitochondrial IRs are common in several stramenopile lineages (Hudspeth et al. 1983; Brien et al. 2014). The function of large IRs in organelle genomes is unclear. In chloroplasts, they sometimes appear to promote genomic stability by reducing mutation rates (Palmer and Thompson 1982; Maier et al. 1995), but in other studies this effect was not observed (Blazier et al. 2016). HR-driven repair of DSBs within the IR using the alternate copy as a template is thought to contribute to the reduced mutation rate observed in previously mentioned studies (Gualberto et al. 2014; Zhu et al. 2016). The importance of HR-mediated repair of DSBs is well established in many organisms (Shrivastav et al. 2008). In *Candida albicans*, the mitochondrial IR is centered on a replication origin, and appears to regulate genome replication through recombination-driven replication with the IR as an initiation site (Gerhold et al. 2010).

In this article, we present new mitochondrial genome sequences of 12 basidiomycete species, and explore the occurrence of the IR and G4DNA motifs in these genomes. Given that both structural phenomena have been associated with genome replication and (in)stability, we hypothesize that the IR is correlated with enriched G4DNA in *Termitomyces*, and that selection for increased genomic stability may have driven expansion of the IR in these species.

Materials and Methods

DNA Material and Sequencing

We selected species for this study based on their phylogenetic position (Hofstetter et al. 2014) and ecology. *Termitomyces* strains were selected to include at least two strains for each of the three most species-rich associated termite host genera (*Macrotermes*: T132, T123, DKA19; *Microtermes*: Mi166, T13; *Odontotermes*: T32, T159) (Aanen et al. 2002). *Termitomyces* strains associated with these genera cover the phylogenetic diversity of this fungal genus. *Blastosporella zonata* and *T. rancida* were selected as representatives of the sister clade of *Termitomyces*. *Myochromella boudieri* was selected as representative of the sister clade to the previous two clades. Finally, we selected *Tricholomella constricta* and *Ast. parasitica* as representatives of more distantly related Lyophyllaceae. Material used for DNA isolation was obtained

from pure cultures. Fungal isolates were grown on malt yeast extract agar (MYA; per liter demi water: 20 g malt extract, 2 g yeast extract, 15 g agar). The *Termitomyces* isolates were grown at 25 °C in the dark. *Tricholomella constricta*, *Ast. parasitica*, *T. rancida*, and *M. boudieri* isolates were grown at 15 °C in the dark. For *Termitomyces* sp. DKA19, no culture was present and hyphal nodules collected from a termite mound were used instead. Nodules were stored in pure ethanol at –20 °C. Prior to extraction, five nodules were rinsed with clean pure ethanol and dried on filter paper.

DNA was extracted using a cetyltrimethylammonium bromide (CTAB) protocol ([supplementary data 1, Supplementary Material](#) online). Library preparation and whole-genome sequencing was performed by Novogene (Hong Kong) using the Illumina HiSeq 2500 platform. Generated reads were 150 bp long and the insert size was 500 bp. Because of the high coverage of mtDNA, we generally used a subsample of one to two million reads to assemble the mitochondrial genomes. This speeds up the assembly process and limits interference of nuclear DNA.

Sequence Assembly

To reconstruct mitochondrial genome sequences from whole-genome sequencing data, we took a reference-based iterative read baiting approach using the IOGA-pipeline (Iterative Organelle Genome Assembly; Bakker et al. 2016). IOGA performs quality filtering and adapter trimming of reads using BBduk, reference mapping using BBmap, and read assembly using both SOAPdenovo2 (Lam et al. 2013) and SPAdes (v3.1.1; Bankevich et al. 2012). Resulting assemblies are evaluated using maximum likelihood with the program ALE (Clark et al. 2013) to assist in identifying the “best” assembly. As IOGA-assemblies did not always capture the full mitochondrial genome sequence in one contig, subsequent scaffolding of contigs was performed if necessary using SSPACE and GapFiller (Boetzer et al. 2011). Final assembly of remaining contigs was performed manually to resolve the IR as follows: 1) contigs covering the IR were identified by a 2-fold increase in sequence coverage; 2) IR-SC boundaries were identified by 95-bp sequence overlaps between the IR and SC contigs; 3) such contigs were joined and the overlapping sequence was merged; 4) reads were mapped back to the resulting scaffold to check whether paired reads surrounding the IR-SC boundary mapped correctly. Some gaps in the assembly arose from sequence artifacts due to long (10 bp+) stretches of monomer repeats. Reads covering these monomers would accumulate sequence errors downstream of the repeat, resulting in termination of the contig. We resolved these gaps by removing the sequence between the gap and the repeat, leaving roughly ten bases of the monomeric repeat intact. We then mapped reads back to the repeat to roughly estimate its length and check if read pairs mapped correctly to either side of the repeat. We identified potential assembly errors by mapping reads back to the draft assembly and inspecting

the SAM-file for regions with reduced coverage, SNPs, or incorrect read pairings.

Annotation

Initial annotation of the first mitochondrial genome (T132) was done manually by performing a BLAST search of the genome against closely related genomes such as *Tricholoma matsutake* and *Pleurotus ostreatus* in the NCBI database. Subsequent annotations used the annotated T132 genome as a reference, by transferring annotations from T132 with a 70% sequence similarity threshold in Geneious v11 (Kearse et al. 2012), and then manually adjusting as needed. Exon boundaries were approximated by comparing amino-acid sequences to references with BlastP. Additional (nonconserved) ORFs were identified with Geneious. We used RNAWeasel (Lang et al. 2007) to identify tRNAs and the small ribosomal subunit. Transcriptomics data (SRA accession SRR5944782) were available for one *Termitomyces* strain of the same suspected species as *Termitomyces* sp. T132 (da Costa et al. 2018), which we used to check our annotation. We aligned RNAseq reads to the annotated genome using TopHat (Trapnell et al. 2009) with Bowtie2 (Langmead and Salzberg 2012) using the sensitive parameter and otherwise default settings. Although for some genes read coverage was limited, for most genes, we confirmed the intron–exon boundaries in our annotation. Images of the annotated genomes were produced using OGDRAW (Lohse et al. 2013).

PCR Confirmation of IR Borders

To confirm the presence of the mitochondrial IR in vitro, a touchdown PCR was conducted. The primers that were used were designed manually based on the mitochondrial genome assembly ([supplementary data 2A, Supplementary Material](#) online). We designed primers to cover all four expected border regions between the IR, SC1, and SC2 for two *Termitomyces* species: T132 and T13. The master mix and PCR program used can be found in the [supplementary data \(2B\), Supplementary Material](#) online. The PCR products were sequenced using the forward primer at Eurofins Genomics (Ebersberg, Germany).

Phylogenetic Analysis of Fungal mtDNA

We reconstructed two phylogenetic trees: one focusing on the Lyophyllaceae, based on 14 core mitochondrial genes; and one covering a representative group of all fungi (James et al. 2006) based on a subset of five mitochondrial genes (*cox1*, *cox2*, *cox3*, *cob*, and *atp6*). We used RevTrans (Wernersson and Pedersen 2003) with default settings to align nucleotide sequences of each gene while maintaining codon structure. Alignments were concatenated and partitioned by PartitionFinder (Lanfear et al. 2012) according to first, second, and third codon position. We performed model

selection and phylogenetic reconstruction using the IQ-TREE program with 5,000 bootstraps and otherwise default settings (Nguyen et al. 2015). For the fungal phylogeny, we included a topological constraint tree (supplementary data 6, Supplementary Material online) to resolve deep divergences conformant to James et al. (2006). We also constrained the Lyophyllaceae clade to conform to our 14 gene analysis, as the five genes contain less phylogenetic signal to accurately resolve these closely related species.

We also ran a MrBayes v3.2.6 analysis (Ronquist et al. 2012) through the Cipes web server (Miller et al. 2010) for the Lyophyllaceae phylogeny with the following parameter settings: 50 million generations, sample frequency 5,000, nst=mixed, four chains, and a burnin percentage of 35. We used a Gamma model of sequence evolution with invariant sites. The topology used in this article was derived from the IQTree Maximum Likelihood analysis and since there were no supported topological conflicts with the Bayesian tree, we include the posterior probability support values from the MrBayes analysis with the ML bootstraps in order to indicate nodal support. We calculated silent and nonsilent substitution rates of mitochondrial genes for the Lyophyllaceae genomes using the R function *kaks* (package *seqinr*), using *Tricholoma matsutake* as outgroup species.

Whole-genome alignment was performed using progressiveMauve v.20150226, (Darling et al. 2004). Species with an IR had one copy of the IR removed to facilitate alignment of duplicated regions.

Detection of G-Quadruplex Sequence Motifs

We used G4Hunter (Bedrat et al. 2016) to scan mtDNA sequences for putative regions capable of G-quadruplex formation. We used strict and relaxed settings (respectively, $w=25$, $s=1.7$, and $w=25$, $s=1.2$) to account for the inherent uncertainty of in silico detection of G4DNA. Because G4Hunter occasionally reports sequences with significant overlap, we eliminated sequences that overlapped a previously reported sequence by >50% of their length. This was done to avoid counting the same potential G-quadruplex twice or more.

GC Skew Analysis

To identify potential locations of replication origins, we created cumulative GC skew graphs (Grigoriev 1998) of our mtDNA assemblies. We used DAMBE v.6.4.29 (Xia and Xie 2001) with default settings to generate GC skew data and integrated the results to obtain the cumulative GC skew.

Results

Overview of Mitochondrial Genomes

We assembled the complete mitochondrial genome for 12 species belonging to the Lyophyllaceae: seven unnamed

species of *Termitomyces* covering the known diversity of the genus with strains from the three known genera of fungus-growing termites *Macrotermes*, *Microtermes*, and *Odontotermes* (Aanen et al. 2002; van de Peppel J.J., M. Nieuwenhuis, D.K. Aanen, Z.W. De Beer unpublished data), *B. zonata*, *T. rancida*, *M. boudieri*, *Tricholomella constricta*, and *Ast. parasitica*. All genomes contained a core set of protein-coding genes (table 1) ubiquitous for fungal mitochondria: *atp6*, *atp8*, *atp9*, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, and *rps3*; as well as 16s and 23s ribosomal subunits, and around 24 tRNAs (not including duplicates, precise number varies per species). The mtDNA of *B. zonata*, *T. rancida*, and all *Termitomyces* specimens featured a large IR region and therefore duplication of all genes contained within, though the gene content of the IR varied from species to species. All species except *Ast. parasitica* had introns in one or more mitochondrial genes. In particular, *cox1* often featured several introns, most of which harbor LAGLIDADG or GIY-YIG endonucleases, a phenomenon commonly found in basidiomycete fungi (Lang et al. 2007).

We found four distinct types of autonomous linear mitochondrial plasmids, two in each of two *Termitomyces* strains (T132 and T123). Such plasmids are commonly found in fungal mitochondria, for example, maranhar/kalilo in *Neurospora* (Chan et al. 1991; Court and Bertrand 1992), and are capable of (partial) insertion within mtDNA. They were presumably assembled by IOGA due to partial shared homology of the plasmid sequence and regions of mtDNA that derive from plasmid insertions. We could distinguish autonomous plasmid sequences from inserted ones using several key features: autonomous plasmids were placed by the assembler in a separate contig whereas inserts were merged with other mtDNA sequences; autonomous plasmids were assembled completely including both terminal IRs (TIRs) and both intact DNA and RNA polymerases, while inserts were generally degenerate and had lost either the TIRs and/or the polymerases; autonomous plasmids showed different (~2-fold higher) coverage compared with mtDNA; and finally, reads mapping to autonomous plasmids showed no paired reads mapping to mtDNA sequences.

When BLASTing plasmid sequences against our mtDNA assemblies, we found significant *E*-value ($\ll 1e-20$) matches with the host mtDNA, as well as occasional matches with mtDNA of other species, suggesting these plasmids were present in the common ancestor of other strains as well, and inserted parts of their DNA in the host mtDNA.

Mitochondrial IR

We detected the mitochondrial IR in our assemblies by a 2-fold increase in read coverage in the duplicated region, due to the assembler merging the identical copies. The inverted nature of the duplication was suggested by inverted read

Table 1

Overview of Mitochondrial Genomes of 12 Lyophyllaceae Species Assembled and Annotated for This Study

| Strain | GenBank ID | mtDNA Size (bp) | IR Size (bp) | GC Content (%) | Introns | | | | | | | | | | |
|---------------------------------|------------|--------------------|-----------------|----------------|---------|-----|------|------|------|------|------|------|-----|-----|---|
| | | | | | atp9 | cob | cox1 | cox2 | nad1 | nad2 | nad4 | nad5 | rns | rnl | |
| <i>Asterophora parasitica</i> | MH725791 | 43,328 | 0 | 31.7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| <i>Tricholomella constricta</i> | MH725800 | 65,087 | 0 | 25.3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| <i>Myochromella boudieri</i> | MH725793 | 99,774 | 0 | 27.9 | 0 | 3 | 9 | 2 | 2 | 0 | 0 | 1 | 0 | 0 | 4 |
| <i>Blastosporella zonata</i> | MH725792 | 200,401 | 56,326 | 34.8 | 1 | 6 | 6 | 3 | 0 | 0 | 1 | 2 | 2 | 2 | 4 |
| <i>Tephroclybe rancida</i> | MH725794 | 126,794 | 33,754 | 38.3 | 0 | 2 | 3 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| <i>Termitomyces</i> sp. Mi166 | MH725795 | 239,317 | 71,805 | 37.8 | 1 | 4 | 15 | 1 | 1 | 1 | 0 | 0 | 2 | 2 | 8 |
| <i>Termitomyces</i> sp. T159 | MH725799 | 157,156 | 41,609 | 39.3 | 1 | 2 | 4 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 3 |
| <i>Termitomyces</i> sp. T32 | MH725797 | 131,333 | 35,619 | 40.9 | 0 | 0 | 5 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| <i>Termitomyces</i> sp. T13 | MH725796 | 155,430 | 44,175 | 34.5 | 0 | 3 | 6 | 2 | 0 | 0 | 0 | 0 | 2 | 2 | 2 |
| <i>Termitomyces</i> sp. DKA19 | MH743217 | 105,724 | 4,463 | 31 | 1 | 2 | 8 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| <i>Termitomyces</i> sp. T123 | MH725798 | 124,711 | 28,228 | 33.1 | 0 | 1 | 6 | 2 | 0 | 1 | 0 | 1 | 0 | 0 | 3 |
| <i>Termitomyces</i> sp. T132 | MG783568 | 146,712 | 37,211 | 31.3 | 1 | 1 | 8 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |

pairings between the ends of the assembled sequence and the borders of the duplicated region. To confirm the presence of the IR, we designed primers around the predicted border regions of the IR and the single copy regions. We succeeded in amplifying products for all four border regions in two strains of *Termitomyces* (T13 and T132, [supplementary data 2C](#), [Supplementary Material](#) online). We took this result as confirmation of the IR hypothesis in all samples with a similar coverage peak. We then copied and pasted the IR sequence in its inverted position to complete each genome sequence. We defined the single copy regions as single copy region 1 (SC1, which for all species barring *B. zonata* contains *cox1*) and single copy region 2 (SC2, which always contains *cox3* and *rps3* at its edges).

The gene content of the IR varies somewhat per species, but in most cases it encompasses the small and large ribosomal subunits, *atp6*, *nad2*, *nad3*, *nad4*, *cox2*, and several tRNAs ([table 2](#)). The border regions with SC2 appear relatively stable in terms of gene content, with *cox3* and *rps3* always just on the outside of the IR. In contrast, border regions surrounding SC1 are highly variable. In *Termitomyces* sp. T132, the 23 s ribosomal subunit overlaps with the IR border, resulting in an incomplete copy of the subunit on one side of the IR. In *T. rancida*, *nad4* is partially duplicated in the same vein.

We compared silent substitution rate estimates of genes of the 12 mitochondrial genomes we sequenced in this study, to test whether genes inside the IR had lowered substitution rates. We found no systematic difference in substitution rates for IR-contained genes between species with and without an IR ([supplementary data 3](#), [Supplementary Material](#) online). K_a/K_s estimates were all <1 which suggests that all genes are under purifying selection as is typical for mtDNA (Soares et al. 2013).

To see if the IR harbored the replication origins as was found for *C. albicans* (Gerhold et al. 2010), we analyzed the cumulative GC skew profile for each mitochondrial genome

with an IR ([supplementary data 5](#), [Supplementary Material](#) online). We were unable to conclusively determine the locations of origins of replications from these graphs, as there was too much variation between profiles, presumably due to inversions and other rearrangements.

Phylogenetic Analysis/Mauve

Our phylogenetic analysis ([fig. 1](#)) shows strong concordance with previous research; the monophyly of *Termitomyces* (Aanen et al. 2002; Frøslev et al. 2003), *B. zonata* and *T. rancida* as the sister clade, the position of *M. boudieri* outside the *T. rancida* clade and the monophyly of *Ast. parasitica* and *T. constricta* (Hofstetter et al. 2014; Bellanger et al. 2015). Our phylogenetic reconstruction shows that the IR has a single origin probably in the common ancestor of the *Termitomyces* clade and its sister clade containing *T. rancida* and *B. zonata*. *Myochromella boudieri* is currently the most closely related taxon to these clades without the IR. We did not find any subsequent losses of the IR in this study, although the span of the IR has significantly decreased in *Termitomyces* sp. DKA19. The three *Termitomyces* spp. associated with *Macrotermes* termites are monophyletic which is consistent with previous studies (Aanen et al. 2002; Nobre et al. 2011). The symbionts associated with *Microtermes* termites show paraphyly, with sp. Mi166 being early branching from the other *Termitomyces* species. Nobre et al. (2011) found a similar occurrence with one *Microtermes* symbiont being placed outside the main clade.

Whole-genome alignment by Mauve ([fig. 2A](#), only four species shown as representation) revealed multiple rearrangement events, including numerous rearrangements among *Termitomyces* species, several of which appear to be linked to contraction/expansion of the IR ([fig. 2B](#)). For example, the change of a number of tRNAs from an upstream position of *cox1* to a downstream position in several *Termitomyces*

Table 2
Gene Content of Inverted Repeat Regions in mtDNA of *Termitomyces* and Related Fungi

| Species | Genes in IR | | | | | | | tRNAs in IR | | | | | | | Genes on Border SC1 | Genes on Border SC2 | | |
|-------------------------------|----------------|-----|------|----------------|----------------|------|------|-------------|-----|-----|-----|-----|-----|-----|---------------------|---------------------|-------------------------------|-----------|
| | rnl | rns | cox2 | nad4 | nad2 | nad3 | atp6 | Ile | Val | Ala | Phe | Met | Thr | His | | | Met (2) | Met (3) |
| <i>Blastosporella zonata</i> | x | x | x | | x | x | x | x | x | x | x | x | | | x | | tRNA-His/nad4 | cox3/rps3 |
| <i>Tephrocye rancida</i> | x | x | x | x ^a | x | x | x | x | x | x | x | x | x | x | x | x | tRNA-Met(4)/nad4 ^a | cox3/rps3 |
| <i>Termitomyces</i> sp. Mi166 | x | x | x | | x | x | x | x | x | x | x | x | x | x | x | | tRNA-Met(3)/nad4 | cox3/rps3 |
| <i>Termitomyces</i> sp. T159 | x | x | x | x | x ^a | | x | x | x | x | x | x | x | x | x | | nad2 ^a /cox1 | cox3/rps3 |
| <i>Termitomyces</i> sp. T32 | x | x | x | | | | x | x | x | x | x | x | x | x | x | | nad4/cox1 | cox3/rps3 |
| <i>Termitomyces</i> sp. T13 | x | x | x | x | x ^a | | x | x | x | x | x | x | x | x | x | | nad2 ^a /tRNA-Gly | cox3/rps3 |
| <i>Termitomyces</i> sp. DKA19 | | | | x ^a | | | | | | | | | x | x | | | tRNA-Ser/nad4 ^a | cox3/rps3 |
| <i>Termitomyces</i> sp. T123 | x ^a | x | x | x | x | x | x | x | x | x | x | x | x | x | x | | cox1/rnl ^a | cox3/rps3 |
| <i>Termitomyces</i> sp. T132 | x ^a | x | x | x | x | x | x | x | x | x | x | x | x | x | x | | cox1/rnl ^a | cox3/rps3 |

NOTE.—Flanking genes for each set of border regions (IR-SC1, IR-SC2) are also shown in this table.
^aGenes partially overlapping with IRs.

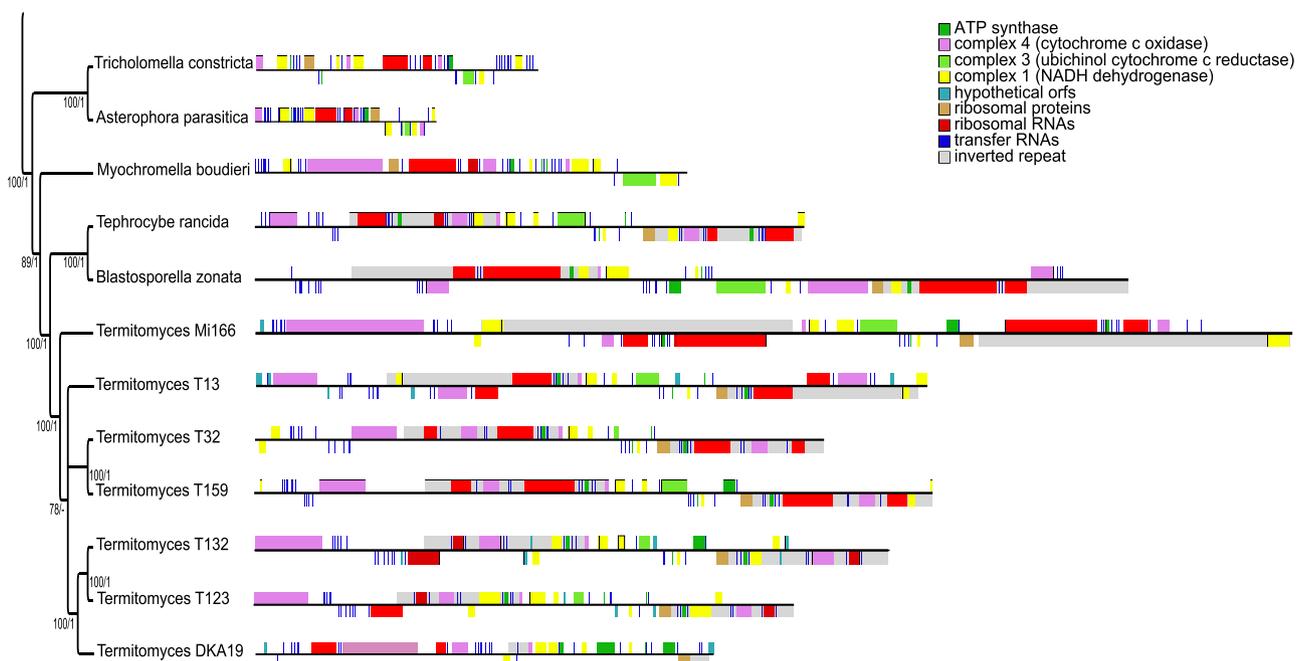


Fig. 1.—Mitochondrial genomes of 12 Lyophyllaceae species. The phylogeny is rooted with *Tricholoma matsutake*, but the genome of this species is not included in the figure. Images of genomes created with OGDRAW (Lohse et al. 2013).

species probably occurred by successive inclusion and expulsion by the IR (Goulding et al. 1996).

G4DNA

We identified G4DNA motifs in most fungal mtDNA under both strict and relaxed settings of G4Hunter. The majority of fungi had comparatively low levels of G4DNA, but some species showed clear peaks of increased G4 content. Both strict and relaxed settings showed similar global patterns of G4DNA content, although for some species, the difference between the two settings was greater than for others (supplemental

data 4, [Supplementary Material](#) online). This may partly be accounted for by a skewness in the false discovery rate, for instance due to differences in GC content or repetitiveness between genomes. Since the overall pattern of strict and relaxed settings was so similar, we will only discuss the results pertaining to the strict analysis here (fig. 3).

G4DNA motif content is a divergent character across fungi, with most species we analyzed showing relatively low frequencies of sequences with G4 potential (<0.5/kb), and several species showing distinctively high G4DNA content. These include the model species *Neurospora crassa*, as well as the basidiomycete *Lentinula edodes*. In addition, all but two

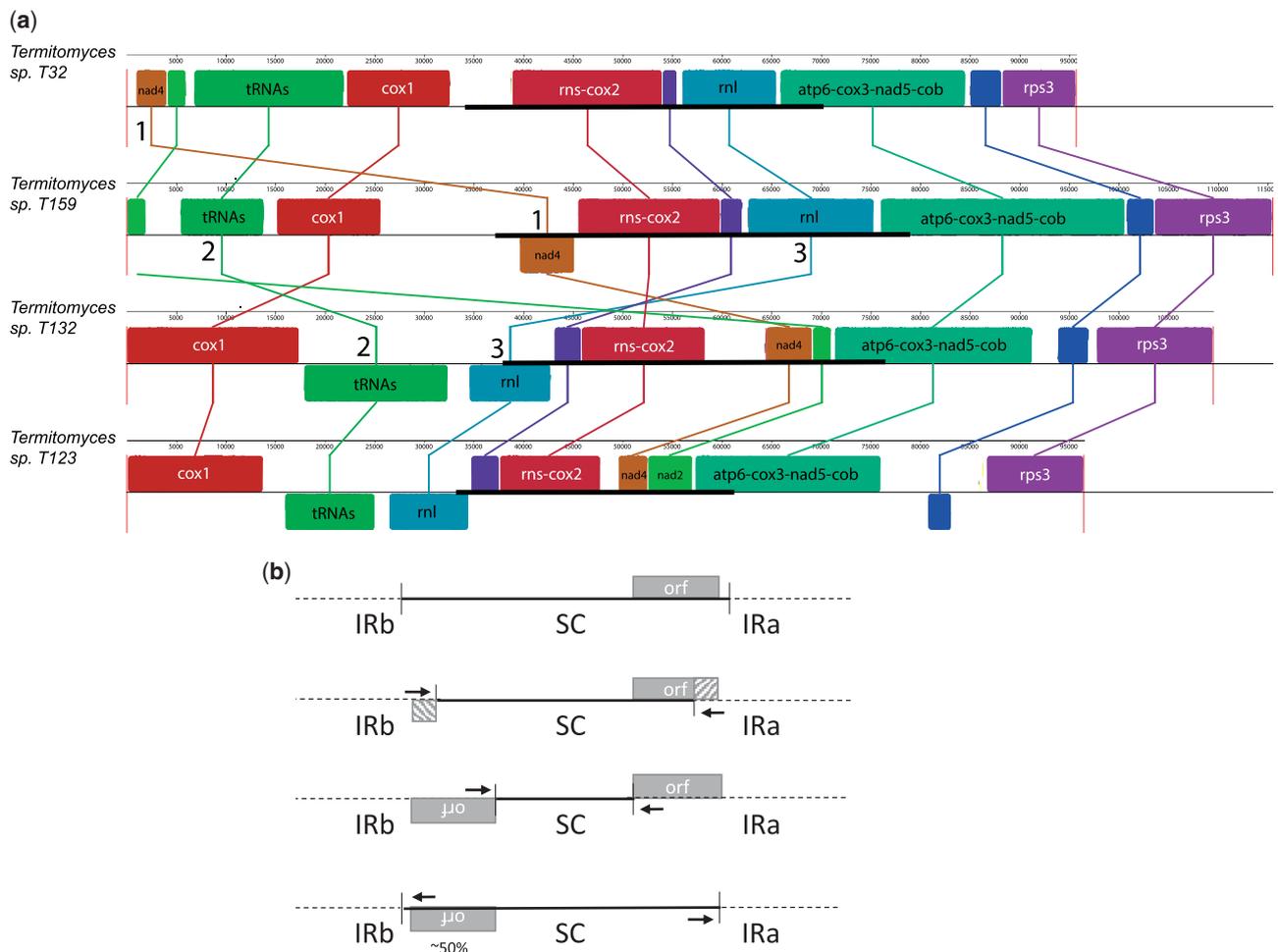


Fig. 2.—(A) Mauve alignment of four *Termitomyces* mitochondrial genomes: sp. T32 (used as reference), T159, T132, and T123. Colored blocks indicate predicted homologies, which are connected through vertical lines. Genes associated with homologous regions are indicated in text for each block. If a block is shown underneath the line it indicates an inversion with respect to the reference. The inverted repeat is represented by a bold line for each genome (only one copy is shown). Several rearrangements are likely due to contraction/expansion of the inverted repeat: 1) The movement of *nad4* from an upstream position of *cox1* in SC1 of sp. T32 to the IR in the other species is probably a result of either contraction or expansion; 2) the repositioning of a large tRNA island from an upstream position of *cox1* in sp. T32 and T159 to a downstream position in sp. T132 and T123 most likely involved both contraction and expansion; 3) the large ribosomal subunit (*rnl*) is only partially included in the IR in sp. T132 and T123, showing a potential expulsion or enveloping in progress. (B) Example of gene translocation through successive inclusion and expulsion by the IR (Goulding et al. 1996). When the IR expands, for instance through illegitimate recombination, it can overlap a flanking ORF. This creates a copy of the overlapping part of the ORF on the opposite IR. The ORF may even end up completely within the IR, in which case two complete copies of the ORF are created. When the IR shrinks and expunges the ORF, one of the copies will disappear, resulting in an ~50% chance of the ORF translocating to a new position.

Termitomyces species, as well as *B. zonata* and *T. rancida* all have estimates $>0.5/\text{kb}$.

The high G4DNA motif content found in our mtDNA assemblies is similar to the high content found in human mtDNA (Bedrat et al. 2016). However, humans like any other vertebrate have miniature, gene-dense mtDNA molecules, with no introns and very little intergenic DNA. In contrast, fungal mtDNA often features numerous introns and long stretches of intergenic DNA. We compared G4DNA motif content of exons, introns, and intergenic DNA of human mtDNA and our fungal mtDNA assemblies, as well as *N. crassa*, to see if G4DNA was located differentially in these regions between

species (fig. 4). We compared observed to expected values assuming unbiased distribution of G4DNA. We found that G4DNA motifs in fungi occurred significantly less in exons of conserved protein-coding genes (*cox1-3*, *cob*, *nad1-6*, *atp6*, *atp8*, *atp9*, and *rps3*), as well as the ribosomal subunits, than expected by chance ($P < 0.05$), whereas in human mtDNA, there was no bias.

To look for a possible relationship between the span of the IR and the frequency of G4DNA motifs, we tested whether G4DNA motifs were more common within the IR than in the SC regions, and found a consistently higher content of G4DNA motifs inside the IR. However, for some species,

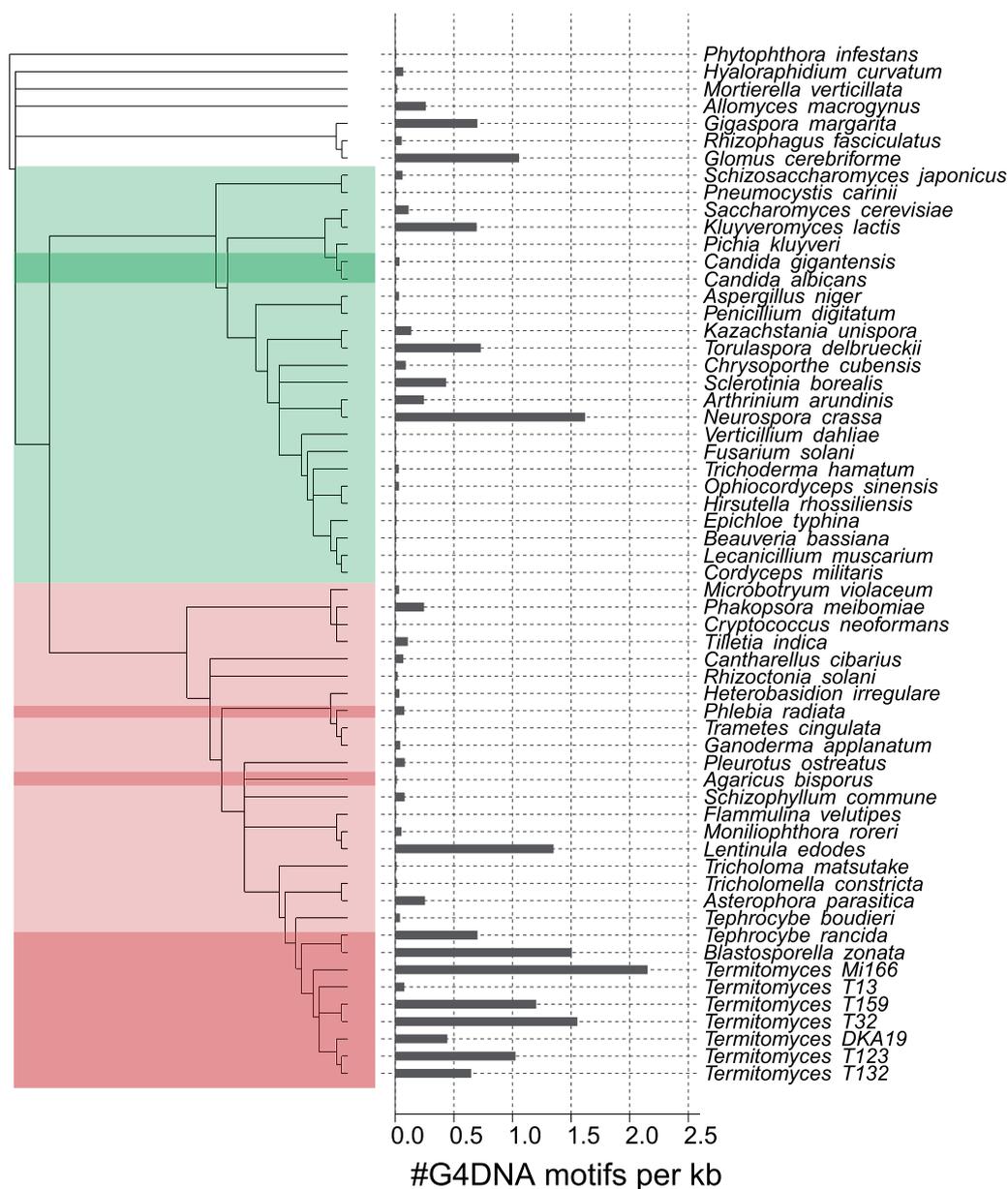


FIG. 3.—Barplot showing average G4DNA motif content per 1-kb mtDNA for 62 species. Values were estimated with strict settings in G4Hunter ($w=25$, $s=1.7$). The phylogenetic tree was reconstructed using Maximum Likelihood with IQtree. Nodes with <70 bootstrap support were collapsed into polytomies. The tree was rooted with *Phytophthora infestans*. Ascomycetes are shaded in green and Basidiomycetes in red. Species with mitochondrial IRs are shaded in a darker hue.

this difference was very small (fig. 5). A phylogenetic paired t -test (phyl.pairedttest, R package “phytools”) showed the difference in G4DNA motif content between IR and SC was statistically significant across species ($P=0.02$). To exclude the possibility that this increase in G4 is simply due to a higher concentration of noncoding DNA in the IR, we calculated what percentage of DNA in the IR and SC regions was part of exons of conserved protein-coding genes or the ribosomal subunits. We found that in all but two cases, the percentage

of exonic DNA was in fact higher in the IR than in the SC regions (table 3).

Discussion

The mitochondrial genomes of *Termitomyces*, *T. rancida*, and *B. zonata* are characterized by a large IR and an enrichment of G4DNA motifs. Although the co-occurrence of these two phenomena could be a neutral effect of hitchhiking, it is

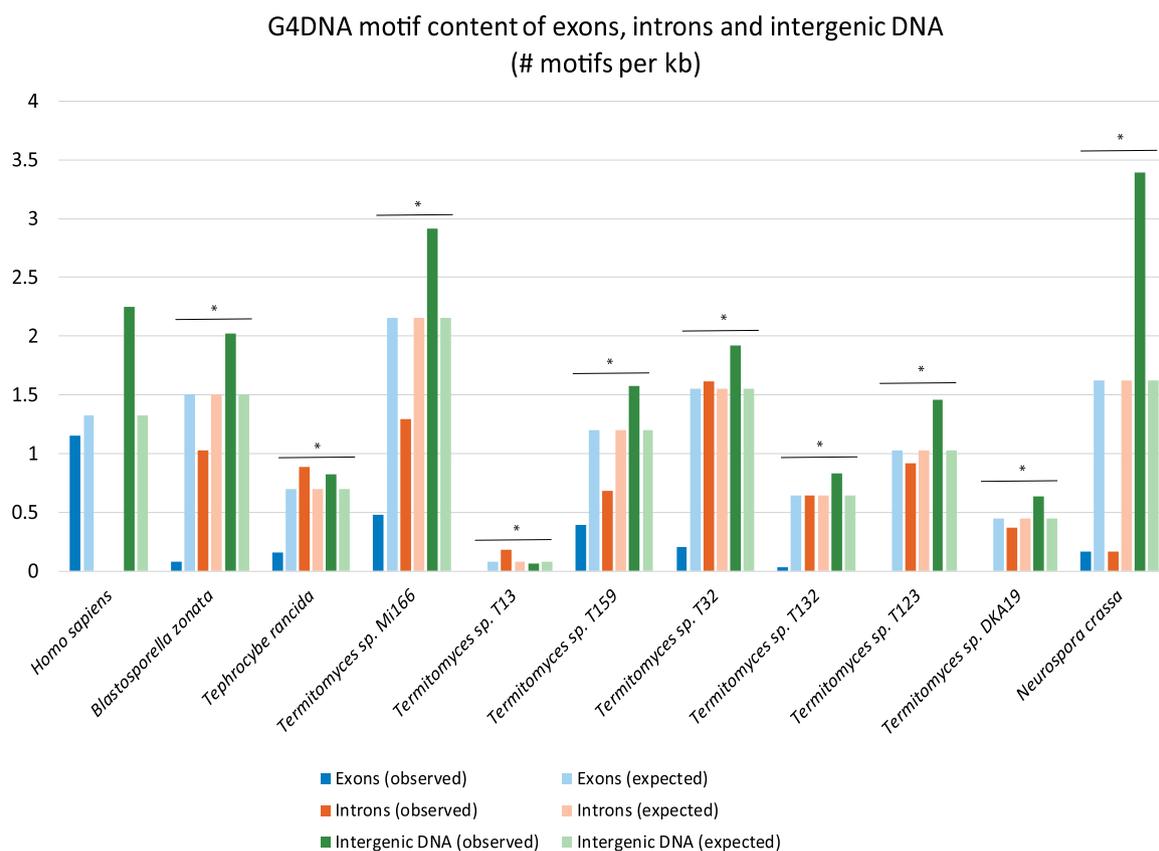


FIG. 4.—Observed and expected G4DNA motif content for exons, introns, and intergenic regions in human and fungal mtDNA. Exons were only considered for the conserved protein-coding genes (*cox1-3*, *nad1-6*, *atp6*, *atp8*, *atp9*, *cob*, and *rps3*) and the ribosomal subunits (*rns* and *rnl*, while these are not protein coding they produce large, functional transcripts). Human mtDNA has no introns. For *Termitomyces sp.* DKA19, T13, and T123, no G4 motifs were observed in coding regions. Asterisks indicate significant deviation of observed values from the expected distribution (χ^2 test, $P < 0.05$).

worth considering whether a functional relation between acquisition of the IR and proliferation of G4DNA exists.

Most G4DNA motifs are located in intergenic regions, among tRNA clusters and surrounding the rRNA subunits. Previous studies on nuclear G4DNA motifs have shown that G4DNA is more often located in promoter regions than expected by chance (Du et al. 2007; Yadav et al. 2008), and therefore may play a regulatory role. Some G4DNA motifs are situated in introns, seemingly embedded in the endonuclease genes. Insertion of GC clusters is a potential method of neutralizing endonuclease activity through frame shifts (Peris et al. 2017). However, most of the intronic ORFs we found remain intact despite the G4DNA. In some species, specifically *N. crassa*, *Termitomyces sp.* Mi166, and *Termitomyces sp.* T159, G4DNA occurs significantly less in introns than expected by chance (fig. 4).

Coding regions in fungal mtDNA almost never appear to contain G4DNA motifs (fig. 4), which suggests that embedding of G4DNA in coding sequences is selected against. This is in contrast to human mtDNA, where G4DNA motifs are found in coding regions close to the expected frequency given

an unbiased distribution. A possible explanation for this is that humans, and likely vertebrates in general, may have more proteins capable of suppressing G4DNA formation in coding regions. A recent study suggests that vertebrates acquired a mechanism to remove G4-prone RNA transcripts (Pietras et al. 2018), possibly reducing the negative posttranscriptional effects of G4DNA motifs in coding regions. Our finding that fungi appear to have strong selection against G4DNA in coding regions suggests they lack such a mechanism.

Generally, G4DNA motif content of fungal mtDNA appears to be low ($<0.5/\text{kb}$) or even zero (fig. 3), but there are some notable exceptions to this trend. Fungi showing high frequencies of G4DNA motifs include the fungus *N. crassa*, a model for genetic and evolutionary studies. The only fungus for which mitochondrial G4DNA motif content was previously reported is *S. cerevisiae* (Capra et al. 2010), and from our analysis, this species appears to have a slightly enriched mitochondrial G4DNA motif content compared with the fungi we currently have data for. Whether G4DNA is a functional part of the genome or accumulates in a neutral fashion, our finding that some species completely lack G4DNA motifs in their mtDNA while in others it is an abundant feature suggests that

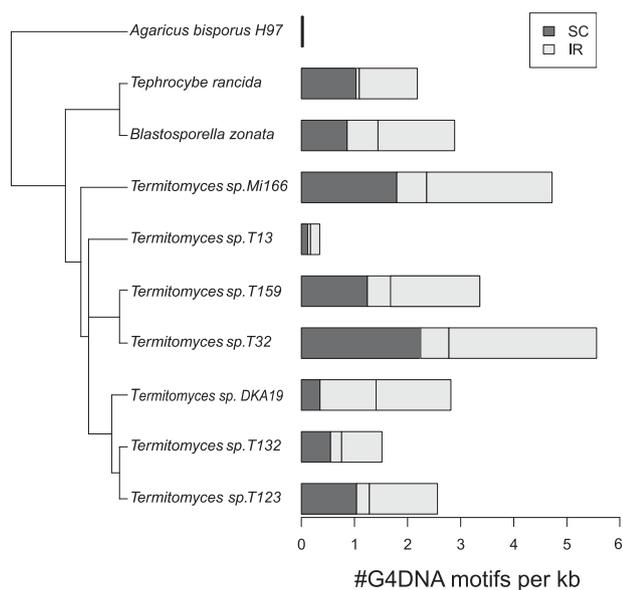


Fig. 5.—Average G4DNA motif content per 1-kb mtDNA for Lyophyllaceae species with IRs, and *Agaricus bisporus* as outgroup. Dark gray denotes the G4DNA motif content for the SC region. Light gray indicates the G4DNA motif content of the IR. The expected 50/50 divide if the G4DNA motif content was equal for IR and SC is shown by a vertical line halfway along each bar. The phylogenetic tree is the same as that of figure 1, but rooted with *A. bisporus* rather than *Tricholoma matsutake* as it is the closest relative to the other species that has a mitochondrial IR.

Table 3

Percentage of DNA in the IR and SC Regions That Belongs to Exons of Conserved Protein-Coding Genes (*cox1-3*, *cob*, *nad1-6*, *atp6*, *atp8*, *atp9* and *rps3*) and the Two Ribosomal Subunits, *rns* and *rnl*.

| Species | Exons (%) | |
|-------------------------------|--------------|--------------|
| | SC | IR |
| <i>Blastosporella zonata</i> | 13.64 | 15.50 |
| <i>Tephroclybe rancida</i> | 18.91 | 30.27 |
| <i>Termitomyces</i> sp. Mi166 | 13.20 | 14.74 |
| <i>Termitomyces</i> sp. T13 | 15.52 | 18.77 |
| <i>Termitomyces</i> sp. T159 | 14.37 | 21.74 |
| <i>Termitomyces</i> sp. T32 | 26.04 | 24.85 |
| <i>Termitomyces</i> sp. T123 | 18.91 | 30.56 |
| <i>Termitomyces</i> sp. T132 | 16.58 | 20.74 |
| <i>Termitomyces</i> sp. DKA19 | 21.16 | 5.18 |

NOTE.—Highest value for each genome is shown in bold.

in both cases mitochondria can function. That G4DNA can have an adverse effect on fitness is clear from its consistent absence from coding regions (fig. 4). As such it seems most likely that in those species where G4DNA is wholly absent selection against G4DNA also affects noncoding regions, or there is limited noncoding space for G4DNA to settle.

Previous studies have reported GC-rich motifs in fungal mtDNA, for example, those surrounding recombination sites in some yeasts (Dieckmann and Gandy 1987;

Liachko et al. 2014), double-hairpin elements (DHEs) in *Allomyces macrogynes* and others (Paquin et al. 2000), and palindromic motifs found in *Neurospora* (Yin et al. 1981). Although not all GC-rich motifs are capable of G-quadruplex formation, we have examined whether some of these sequences qualified as G4DNA motifs in our analysis. Of the 89 GC-rich DHEs reported for *All. macrogynes*, only four overlapped G4DNA motifs. Similarly, more than a hundred GC-rich clusters are reported for the mtDNA of the yeast *S. cerevisiae* (Wolters et al. 2015), while the estimated number of G4DNA motifs was eight. It therefore seems that most of these GC-rich motifs are unlikely to form G4DNA. However, in the case of *N. crassa*, the GC-rich elements reported in Yin et al. generally seem to overlap with G4DNA motifs. That they were not identified as such at the time is understandable considering G4DNA was still relatively obscure and its biological relevance unknown.

Large (1 kb+) mitochondrial IRs, while rare in fungi, occur among others in *Candida*, *Agaricus*, *Phlebia radiata*, and *Agrocybe aegerita*. The IRs found in *Termitomyces*, *T. rancida*, and *B. zonata* are unprecedented in terms of span, taking up half of the genome in most species. While this could be the result of a neutral process of incremental expansion of the IR, the co-occurrence of large amounts of G4DNA motifs provides a possible selective benefit for this increase in size. G4DNA is known to cause frequent DSBs, while IRs can efficiently repair DSBs through HR (Lambert et al. 2010). Consistent with this, G4DNA motifs are slightly but significantly more frequent within the IR than outside of it (fig. 5), which could indicate that selection against G4DNA is weakened in the IR because the deleterious effects of DSBs are mitigated. Although the difference between G4DNA motif content within and outside the IR is only marginal for some species, it is consistent across species. It should be noted that HR can occur outside the IR as well, between different copies of mtDNA. However, the IR can undergo HR even when other copies of mtDNA are absent, and the potential frequency of HR increases with repeat copy number (Fujitani et al. 1995).

The enrichment of G4DNA in the IRs of *Termitomyces* could also be due to the colocalization of replication origins in the repeats. The association of G4DNA with replication origins is supported in other organisms (Paeschke et al. 2011; Valton et al. 2014; Valton and Prioleau 2016), and the potential role of mitochondrial IRs in recombination-driven replication of mtDNA is supported by experimental results obtained from *C. albicans* (Gerhold et al. 2010). In chloroplasts of maize, the IRs function as termini of linear monomers of ptDNA, and harbor the replication origins (Oldenburg and Bendich 2016). Some studies (Grigoriev 1998; Gerhold et al. 2010; Xia 2012) have employed GC-skew analyses to estimate replication origins in mtDNA sequences, however, the reliability of this method is dependent on the replication method and is severely reduced if the

genome is frequently rearranged. We produced GC-skew graphs for our assembled mtDNA sequences but could not conclusively determine replication origin positions from them (supplementary data 5, Supplementary Material online). The IR itself appears to show no reduced mutation rate in contrast to what was observed in some other studies (Palmer and Thompson 1982; Maier et al. 1995) (supplementary data 3, Supplementary Material online). However, HR is evident from the lack of divergence between the two copies of the repeat. Several key genes are duplicated by the IR, but it is unclear whether a dosage effect plays a role in the emergence of organellar IRs, since the role of mitochondrial genes is highly conserved and most species do not have IRs or duplicated genes in their mtDNA.

The IR is unstable between species in terms of span (table 1). The border regions show distinct behaviors: one side (IR-SC1) is very dynamic, and frequent gene rearrangements around this border result from regular expansions and retractions of the IR. Plasmid insertions surrounding the IR-SC1 border region may also indicate instability. However, the IR-SC2 border appears to be highly stable, with *cox3* and *rps3* always flanking this side of the IR (table 2). Significant gene rearrangements that probably involved the IR (fig. 2) suggest the gene content of the IR was more variable in the ancestors of some species.

Termitomyces sp. DKA19 is peculiar due to the greatly reduced size of its IR (table 1). Despite its reduced size the IR is centered on a pocket of locally enriched G4DNA motifs, giving more credence to the hypothesis that the IR is in some way affecting the G4DNA content. However, it remains to be seen whether the decrease in IR span caused a reduction in global G4DNA content or was rather a result from that reduction. Conversely, in *Termitomyces* sp. T13, the number of G4DNA motifs is highly reduced across the entire length of the genome, but the size of the IR has been maintained.

Organellar IRs, including those of plant chloroplasts, stramenopile mtDNA, and all but one reported in this study (*Termitomyces* sp. DKA19), often encompass the ribosomal subunits. This mimics the multitudinous tendency of nuclear ribosomal DNA, and it may be selectively beneficial to harbor the highly conserved rDNA in a region of increased genomic stability. As ribosomal DNA may be associated with G4DNA (Capra et al. 2010), the increased G4DNA content of the IR of *Termitomyces* and relatives therefore could simply be due to the presence of rDNA. However, that would not explain why the IR of *Termitomyces* sp. DKA19, which does not contain ribosomal DNA, has an elevated G4DNA motif content compared with the rest of the genome.

It is unclear how the IR originated. A small mitochondrial IR in *Agr. aegerita* originated as the result of an insertion of a mitochondrial plasmid. Such plasmids and their insertion sites were also detected in several of our genome reconstructions.

The insertion mechanism of these linear plasmids involves an expansion of their TIRs, and such an event may have given rise to the progenitor of the large IR observed in *Termitomyces* and its sister clade.

Conclusion

We report a large IR and an apparent increase in G4DNA content in the mitochondrial genomes of *Termitomyces*, *B. zonata*, and *T. rancida*. We propose two nonmutually exclusive functional explanations for the observed correlation between the two structural phenomena: 1) the IR helps repair DSBs caused by G4DNA; 2) both structures are involved in the replication of mtDNA. We also provide the first comparative analysis of G4DNA content in mtDNA in fungi. The large discrepancy in mitochondrial G4DNA content between fungal species raises an opportunity for examining the effects of G4DNA on genome stability, replication, recombination, and transcription. Such studies would benefit from comparing similar species with different G4DNA content. In addition, some of the fungi with high mitochondrial G4DNA content are already established model species, *N. crassa* in particular. These species would be ideal candidates for experimental studies on G4DNA function and evolution, and could improve our insight in how G4DNA affects our own (mitochondrial) DNA.

Finally, we have shown that in fungal mtDNA coding regions are significantly depleted in G4DNA motifs, in contrast to humans. This suggests that fungi have stronger selection against G4DNA in exons, as they might lack the means to regulate G4DNA formation posttranscriptionally.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Sabine Vreeburg for generating the RNA sequence alignment. We also thank Ben Auxier for helpful comments on the draft version of this article. We also thank three anonymous reviewers for their feedback. This study was funded by NWO-Vici (Grant 86514007).

Literature Cited

- Aanen DK, et al. 2002. The evolution of fungus-growing termites and their mutualistic fungal symbionts. *Proc Natl Acad Sci U S A*. 99(23):14887–14892.
- Aldrich J, Cherney B, Merlin E, Williams C, Mets L. 1985. Recombination within the inverted repeat sequences of the *Chlamydomonas reinhardtii* chloroplast genome produces two orientation isomers. *Curr Genet*. 9(3):233–238.

- Bakker FT, et al. 2016. Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. *Biol J Linn Soc.* 117(1):33–43.
- Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477.
- Bedrat A, Lacroix L, Mergny JL. 2016. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* 44(4):1746–1759.
- Bellanger JM, et al. 2015. Plunging hands into the mushroom jar: a phylogenetic framework for Lyophyllaceae (Agaricales, Basidiomycota). *Genetica* 143(2):169–194.
- Bharti SK, et al. 2014. DNA sequences proximal to human mitochondrial DNA deletion breakpoints prevalent in human disease form G-quadruplexes, a class of DNA structures inefficiently unwound by the mitochondrial replicative twinkle helicase. *J Biol Chem.* 289(43):29975–29993.
- Biffi G, Tannahill D, Miller J, Howat WJ, Balasubramanian S. 2014. Elevated levels of G-quadruplex formation in human stomach and liver cancer tissues. *PLoS One* 9. e102711.
- Birky CW. 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu Rev Genet.* 35(1):125–148.
- Blazier JC, et al. 2016. Variable presence of the inverted repeat and plasmome stability in *Erodium*. *Ann Bot.* 117(7):1209–1220.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27(4):578–579.
- Brien MAO, Misner I, Lane CE. 2014. Mitochondrial genome sequences and comparative genomics of *Achlya hypogyna* and *Thraustotheca clavata*. *J Eukaryot Microbiol.* 61(2):146–154.
- Burger G, Gray MW, Lang BF. 2003. Mitochondrial genomes: anything goes. *Trends Genet.* 19(12):709–716.
- Capra JA, Paeschke K, Singh M, Zakian VA. 2010. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLoS Comput Biol.* 6(9). e1000861.
- Chan BS, Court DA, Vierula PJ, Bertrand H. 1991. The kalilo linear senescence-inducing plasmid of *Neurospora* is an invertron and encodes DNA and RNA polymerases. *Curr Genet.* 20:225–237.
- Clark SC, Egan R, Frazier PI, Wang Z. 2013. ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. *Bioinformatics* 29(4):435–443.
- Court DA, Bertrand H. 1992. Genetic organization and structural features of Maranhar, a senescence-inducing mitochondrial linear plasmid of *Neurospora crassa*. *Curr Genet.* 22(5):385–397.
- Cui Y, Kong D, Ghimire C, Xu C, Mao H. 2016. Mutually exclusive formation of G – quadruplex and i – motif is a general phenomenon governed by steric hindrance in duplex DNA. *Biochemistry* 55(15):2291–2299.
- da Costa RR, et al. 2018. Enzyme activities at different stages of plant biomass decomposition in three species of fungusgrowing termites. *Appl Environ Microbiol.* 84. e01815-17.
- Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14(7):1394–1403.
- Dieckmann CL, Gandy B. 1987. Preferential recombination between GC clusters in yeast mitochondrial DNA. *EMBO J.* 6(13):4197–4203.
- Du Z, Kong P, Gao Y, Li N. 2007. Enrichment of G4 DNA motif in transcriptional regulatory region of chicken genome. *Biochem Biophys Res Commun.* 354(4):1067–1070.
- Du Z, Zhao Y, Li N. 2008. Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res.* 18(3):233–241.
- Eddy J, Maizels N. 2006. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* 34(14):3887–3896.
- Frøslev TG, Aanen DK, Læssøe T, Rosendahl S. 2003. Phylogenetic relationships of *Termitomyces* and related taxa. *Mycol Res.* 107(11):1277–1286.
- Fujitani Y, Yamamoto K, Kobayashi I. 1995. Dependence of frequency of homologous recombination on the homology length. *Genetics* 140(2):797–809.
- Garg R, Aggarwal J, Thakkar B. 2016. Genome-wide discovery of G-quadruplex forming sequences and their functional relevance in plants. *Sci Rep.* 6:31–35.
- Gerhold JM, Aun A, Sedman T, Jöers P, Sedman J. 2010. Strand invasion structures in the inverted repeat of *Candida albicans* mitochondrial DNA reveal a role for homologous recombination in replication. *Mol Cell.* 39(6):851–861.
- Goulding SE, Olmstead RG, Morden CW, Wolfe KH. 1996. Ebb and flow of the chloroplast inverted repeat. *Mol Gen Genet.* 252(1–2):195–206.
- Grigoriev A. 1998. Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res.* 26(10):2286–2290.
- Gualberto JM, et al. 2014. The plant mitochondrial genome: dynamics and maintenance. *Biochimie* 100:107–120.
- Hänsel-Hertsch R, Di Antonio M, Balasubramanian S. 2017. DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat Rev Mol Cell Biol.* 18(5):279–284.
- Hofstetter V, et al. 2014. Taxonomic revision and examination of ecological transitions of the Lyophyllaceae (Basidiomycota, Agaricales) based on a multigene phylogeny. *Cryptogam Mycol.* 35(4):399–425.
- Huang W-C, et al. 2015. Direct evidence of mitochondrial G-quadruplex DNA by using fluorescent anti-cancer agents. *Nucleic Acids Res.* 43:gkv1061.
- Hudspeth MES, Shumard DS, Bradford CJR, Grossman LI. 1983. Organization of *Achlya* mtDNA: a population with two orientations and a large inverted repeat containing the rRNA genes (intramolecular recombination/circular restriction NIAP/oomycete). *Proc Natl Acad Sci U S A.* 80(1):142–146.
- James TY, et al. 2006. Reconstructing the early evolution of fungi using a six-gene phylogeny. *Nature* 443(7113):818–822.
- Johnson JE, Smith JS, Kozak ML, Johnson FB. 2008. In vivo veritas: using yeast to probe the biological functions of G-quadruplexes. *Biochimie* 90(8):1250–1263.
- Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649.
- Koole W, et al. 2014. A polymerase theta-dependent repair pathway suppresses extensive genomic instability at endogenous G4 DNA sites. *Nat Commun.* 5:1–10.
- Lam T-W, et al. 2013. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:1–6.
- Lambert S, et al. 2010. Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange. *Mol Cell.* 39(3):346–359.
- Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol.* 29(6):1695–1701.
- Lang BF, Laforest MJ, Burger G. 2007. Mitochondrial introns: a critical view. *Trends Genet.* 23(3):119–125.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9(4):357–359.
- Liachko I, et al. 2014. GC-rich DNA elements enable replication origin activity in the methylotrophic yeast *Pichia pastoris*. *PLoS Genet.* 10. e1004169.
- Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* 41:575–581.

- Lynch M, Koskella B, Schaack S. 2006. Mutation pressure and the evolution of organelle genomic architecture. *Science* 311(5768):1727–1730.
- Maier RM, Neckermann K, Igloi GL, Ko H. 1995. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. *J Mol Biol.* 251(5):614–628.
- Maizels N. 2006. Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat Struct Mol Biol.* 13(12):1055–1059.
- Mendoza O, Bourdoncle A, Boulé JB, Brosh RM, Mergny JL. 2016. G-quadruplexes and helicases. *Nucleic Acids Res.* 44(5):1989–2006.
- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. *Proc Gateway Comput Environ Work.* 1–8. doi: 10.1109/GCE.2010.5676129
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Nobre T, Koné NA, Konaté S, Linsenmair KE, Aanen DK. 2011. Dating the fungus-growing termites' mutualism shows a mixture between ancient codiversification and recent symbiont dispersal across divergent hosts. *Mol Ecol.* 20(12):2619–2627.
- Nobre T, Koopmanschap B, Baars JJ, Sonnenberg AS, Aanen DK. 2014. The scope for nuclear selection within Termitomyces fungi associated with fungus-growing termites is limited. *BMC Evol Biol.* 14. 24902958.
- Oldenburg DJ, Bendich AJ. 2016. The linear plastid chromosomes of maize: terminal sequences, structures, and implications for DNA replication. *Curr Genet.* 62(2):431–442.
- Paeschke K, Capra JA, Zakian VA. 2011. DNA replication through G-quadruplex motifs is promoted by the *S. cerevisiae* Pif1 DNA helicase. *Cell* 145(5):678–691.
- Palmer JD. 1983. Chloroplast DNA exists in two orientations. *Nature* 301(5895):92–93.
- Palmer JD. 1985. Comparative organization of chloroplast genomes. *Annu Rev Genet.* 19(1):325–354.
- Palmer JD, Thompson WF. 1982. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* 29(2):537–550.
- Paquin B, Laforest MJ, Lang BF. 2000. Double-hairpin elements in the mitochondrial DNA of allomyces: evidence for mobility. *Mol Biol Evol.* 17(11):1760–1768.
- Peris D, et al. 2017. Mitochondrial introgression suggests extensive ancestral hybridization events among *Saccharomyces* species. *Mol Phylogenet Evol.* 108:49–60.
- Pietras Z, et al. 2018. Dedicated surveillance mechanism controls G-quadruplex forming non-coding RNAs in human mitochondria. *Nat Commun.* 9. 2558.
- Rawal P, et al. 2006. Genome-wide prediction of G4 DNA as regulatory motifs: role in *Escherichia coli* global regulation. *Genome Res.* 16(5):644–655.
- Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Softw Syst Evol.* 61:539–542.
- Schaffitzel C, et al. 2001. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Styloynchia lemnae* macro-nuclei. *Proc Natl Acad Sci U S A.* 98(15):8572–8577.
- Shrivastav M, De Haro LP, Nickoloff JA. 2008. Regulation of DNA double-strand break repair pathway choice. *Cell Res.* 18(1):134–147.
- Smith DR, Keeling PJ. 2015. Mitochondrial and plastid genome architecture: reoccurring themes, but significant differences at the extremes. *Proc Natl Acad Sci U S A.* 112(33):10177–10184.
- Soares P, et al. 2013. Evaluating purifying selection in the mitochondrial DNA of various mammalian species. *PLoS One* 8. e58993.
- Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111.
- Turmel M, Otis C, Lemieux C. 1999. The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: insights into the architecture of ancestral chloroplast genomes. *Proc Natl Acad Sci U S A.* 96(18):10248–10253.
- Valton AL, et al. 2014. G4 motifs affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.* 33(7):732–746.
- Valton AL, Prioleau MN. 2016. G-quadruplexes in DNA replication: a problem or a necessity? *Trends Genet.* 32(11):697–706.
- Verma A, Yadav VK, Basundra R, Kumar A, Chowdhury S. 2009. Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cells. *Nucleic Acids Res.* 37(13):4194–4204.
- Wernersson R, Pedersen AG. 2003. RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res.* 31(13):3537–3539.
- Wilson AJ, Xu J. 2012. Mitochondrial inheritance: diverse patterns and mechanisms with an emphasis on fungi. *Mycology* 3:158–166.
- Wolters JF, Chiu K, Fiumera HL. 2015. Population structure of mitochondrial genomes in *Saccharomyces cerevisiae*. *BMC Genomics* 16:1–13.
- Xia X, Xie Z. 2001. DAMBE: software package for data analysis in molecular biology and evolution. *J Hered.* 92(4):371–373.
- Xia X. 2012. DNA replication and strand asymmetry in prokaryotic and mitochondrial genomes. *Curr Genomics.* 13(1):16–27.
- Xu J, Zhang Y, Pun N. 2013. Mitochondrial recombination in natural populations of the button mushroom *Agaricus bisporus*. *Fungal Genet Biol.* 55:92–97.
- Yadav VK, Abraham JK, Mani P, Kulshrestha R, Chowdhury S. 2008. QuadBase: genome-wide database of G4 DNA – occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. *Nucleic Acids Res.* 36:381–385.
- Yin S, Heckman J, RajBhandary UL. 1981. Highly conserved GC-rich palindromic DNA sequences flank tRNA genes in *Neurospora crassa* mitochondria. *Cell* 26(3):325–332.
- Zaug AJ, Podell ER, Cech TR. 2005. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc Natl Acad Sci U S A.* 102(31):10864–10869.
- Zhu A, Guo W, Gupta S, Fan W, Mower JP. 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol.* 209(4):1747–1756.

Associate editor: Sandra Baldauf