Evolution of mutualisms in the basidiomycete genus *Termitomyces*

Mathijs Nieuwenhuis

Evolution of mutualisms in the basidiomycete genus *Termitomyces*

Mathijs Nieuwenhuis

Propositions

1. Biological data archives such as GenBank need to be commercialized in order to increase their user-friendliness. (this thesis)
2. Phylogenomic reconstructions are often compromised by the volume of their underlying data. (this thesis)
3. Evolution should not generally be referred to as a theory.
4. A successful high school science education imparts an interest in science primarily and scientific knowledge secondarily.
5. Decolonization of scientific knowledge to account for other epistemologies is unwarranted.
6. Dutch high school students should be educated on psychology at the expense of French.
7. Schools and universities should not be formally associated with any religion, be it through their name, curriculum, activities, funding, or any other aspect.
8. Healthcare workers and teachers would have more political leverage if they drove tractors to work.

Propositions belonging to the thesis, entitled:

The evolution of mutualisms in the basidiomycete genus *Termitomyces*

Mathijs Nieuwenhuis

Wageningen, 19 October 2021

Evolution of mutualisms in the basidiomycete genus *Termitomyces*

Mathijs Nieuwenhuis

Thesis Committee

Promotors

Prof. Dr B. J. Zwaan
Professor of Genetics Wageningen University & Research

Prof. Dr D.K. Aanen
Personal chair at the Laboratory of Genetics
Wageningen University & Research

Co-promotor

Dr F.T. Bakker
Assistant Professor, Biosystematics
Wageningen University & Research

Other members

Prof. Dr T.J.G. Ettema, Wageningen University & Research
Prof. Dr J.I. Korb, University of Freiburg, Germany
Prof. Dr H. Johannesson, Uppsala University, Sweden
Prof. Dr H.A.B. Wösten, Utrecht University

Evolution of mutualisms in the basidiomycete genus *Termitomyces*

Mathijs Nieuwenhuis

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public on Tuesday 19 October 2021
at 1.30 p.m. in the Aula.

Mathijs Nieuwenhuis
Evolution of mutualisms in the basidiomycete genus *Termitomyces*
PhD thesis, Wageningen University, Wageningen, the Netherlands (2021)
With references, with summary in English
ISBN: 978-94-6395-916-2
DOI: https://doi.org/10.18174/550785

"I do not believe that we know anything for certain but everything to different degrees of probability"

Christiaan Huygens (1629-1695)

Table of contents

CHAPTER 1

1

General introduction

Cooperation is the foundation of life, as molecules, genes, cells and organisms band together to form more complex entities. Principles that govern the evolution of cooperation in biology are also likely to apply to our own cooperative systems, including agriculture and economics. Important problems for cooperative groups include selfish interests conflicting with the common interest, and individuals cheating the system. Despite these apparent problems long-term cooperative systems are highly successful in both biology and human society. This paradox has long drawn evolutionary biologists to study the origins and mechanisms of biological cooperation (Maynard Smith & Szathmáry 1997; Nowak 2006; Axelrod & Hamilton 1981; Maynard Smith 1964; Hamilton 1964; Fisher 1930; Haldane 1932).

In this thesis I study the origin, mechanism and stability of biological cooperation at three different levels of selection in the symbiotic fungus *Termitomyces*. The result of an ancient domestication process, *Termitomyces* fungi are dependent on fungus-farming termites for their survival and reproduction. This fungus-termite mutualism is the first level of biological cooperation I explore to uncover the ancestral conditions from which it evolved. The second level concerns the ancient mutualism between eukaryote nuclei and mitochondria. As sexual reproduction in *Termitomyces* potentially involves biparental transmission of mitochondria, I investigate whether heteroplasmy and recombination of mitochondrial DNA occurs, as these phenomena could destabilize the cooperation between nuclear and mitochondrial genomes. Finally, as is the case for many fungi, the mitochondria of *Termitomyces* species are frequent hosts to mitochondrial plasmids. These plasmids are known to exchange their DNA with that of a host mitochondrial genome, but their effect on host fitness is unclear. I analyse the DNA of a wide range of plasmids retrieved from *Termitomyces* and related fungi and propose a potential mutualistic interaction through genetic addiction of mitochondria to plasmids in some species. This constitutes the third level of biological interaction studied in this thesis. For biological complexity to evolve, functional cooperation is required at each level of selection, starting at the lowest. Combined, the papers in this thesis examine the evolution of cooperation from one of the lowest levels of biological complexity (the level of genes and genomes) to one of the highest (interspecific mutualisms) in the symbiotic fungus *Termitomyces* (Figure 1).

Major transitions in evolution

The evolution of life is marked by numerous transitions from solitary to cooperative states, termed by John Maynard Smith and Eörs Szathmáry as 'major transitions' (Smith & Szathmary 1997). Some examples are the evolution of multicellularity, the evolution of sex, eukaryogenesis, and (mutualistic) symbioses. Such transitions can be said to increase

the complexity of a life form, since they add an additional hierarchical tier for selection to act on. One of the ongoing challenges in evolutionary biology is understanding how these major transitions in biological complexity originate and persist, as selection can act in opposing directions on different tiers of such hierarchical systems.

A famous example to illustrate this problem is the tragedy of the commons, a hypothetical scenario outlined by Garrett Hardin (Hardin 1968) based on a pamphlet by William Forster Loyd from 1833. It describes a community of shepherds that share a common pasture for their sheep to graze on. The pasture can only support a limited number of sheep; therefore, each shepherd should restrict the size of his herd to accommodate the others and prevent overgrazing. The 'tragedy' occurs when a shepherd realizes he can cheat slightly and add one more sheep to his herd, the benefit of which he enjoys alone while the cost in additional grazing is shared among all the others. As more and more shepherds start to cheat in this way the pasture becomes overgrazed and the communal system collapses: any individual benefit is gone and everyone loses. The interests of the individuals (add more sheep) were opposed to the interests of the group (prevent overgrazing).

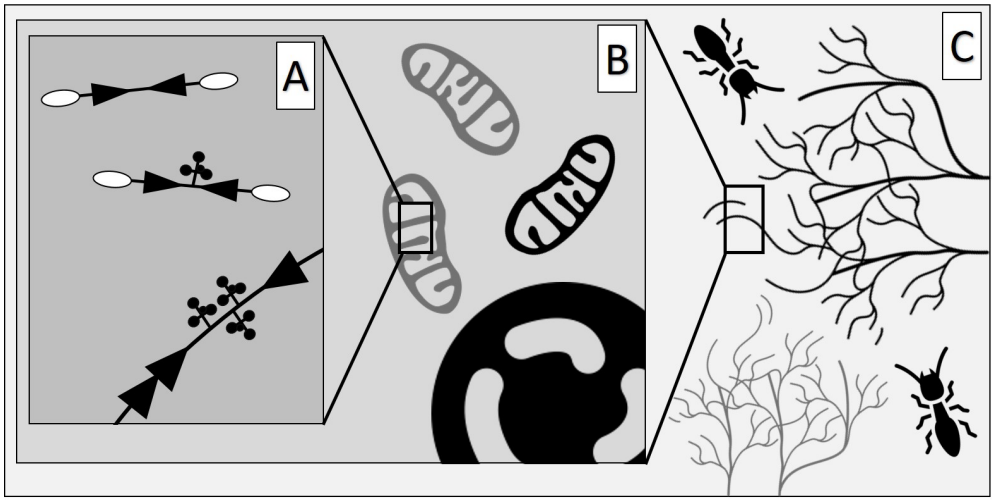


Figure 1: Levels of biological complexity. Three levels of selection at which cooperation between self-replicating entities has evolved in *Termitomyces*: A) Linear mitochondrial plasmids are autonomous genetic elements found in the mitochondria of various fungi and plants. Transfer of mitochondrial tRNA genes to plasmids may result in genetic addiction and mutualistic co-dependency between a plasmid and mitochondrion. B) The mutualism between precursor eukaryote cells and mitochondria evolved over two billion years ago. Genetic variation of mitochondria between and within cells is usually restricted, presumably to avoid selection of selfish variants. C) The mutualism between *Termitomyces* and the Macrotermitinae is one of the few known cases of non-human domestication.

Mutualisms

Charles Darwin stated that his theory of evolution by natural selection could be disproven by showing a species possessed a trait for the exclusive benefit of another species (Darwin 1859). This statement could raise the question whether the production of alluring fruits and nectar-filled flowers by plants is the disproof of evolution that creationists have sought for decades. However, the mutualistic interactions we observe between species always confer a fitness benefit to both the donor and the recipient. In the case of fruits and flowers, for instance, the cost incurred in their production to the plant is offset by the increased dispersal of its seeds and pollen, respectively.

Despite their reciprocal nature mutualisms are driven by self-interest (Leigh Jr 2010). As the parable of the tragedy of the commons predicts, mutualistic symbionts tend to try to minimize the costs of their partnership while maximizing the gains (Rankin et al. 2007; Trivers 1971). If left unchecked, this tendency will result in the breakdown of mutualisms as symbionts become "cheaters": individuals that benefit from the interaction without contributing to it. Examples of mechanisms to avoid this breakdown are restricting genetic diversity among symbionts (resulting among other effects in increased kin selection), sanctioning, or offering diminishing returns (e.g., in resource exchanges) (Rankin et al. 2007; Leigh Jr 2010) (Table 1). In many mutualisms, multiple mechanisms are in place to resolve conflicts between symbionts. Mitochondria, for example, are governed in many eukaryotes by both restricting genetic diversity through uniparental inheritance and other bottlenecks (Christie et al. 2015), and sanctioning by selective mitophagy (Kim et al. 2007).

Three chapters in this thesis (Chapter 3, Chapter 4 and Chapter 5) feature research related to the origin and maintenance of various kinds of mutualistic interactions, all associated with the basidiomycete *Termitomyces*: a phylogenomic reconstruction of the domestication of *Termitomyces* by termites (Chapter 3), a discovery of potential genetic addiction of mitochondria to mitochondrial plasmids (Chapter 4), and an analysis of mitochondrial inheritance and recombination (Chapter 5). Furthermore, Chapter 2 comprises a comparative analysis of complete mitochondrial genomes of *Termitomyces* and related fungi, which laid the foundation for the research reported in the other three Chapters. Combined, the research in this thesis illustrates how mutualisms can arise at the level of the organism and the level of the gene. *Termitomyces* serves as a good system to study the origin and stability of mutualisms, as it represents a unique evolutionary trajectory of domestication that has lasted at least 30 million years (Aanen et al. 2002). It serves as an independent comparison to similar domestications such as those found in fungal cultivars of the attine ants (Mueller et al. 2001) and ambrosia beetles (Cassar & Blackwell 1996), and even human agriculture.

Table 1: Mechanisms to maintain stability and prevent cheating in host-symbiont relations

Mechanism	Examples
Restrict genetic diversity	Uniparental inheritance of endosymbionts (Birky 2001; Koga et al. 2012; Enticknap et al. 2006)
Sanctioning	Bio-luminescent bacteria and squid (Visick et al. 2000), plant rhizobia mutualisms (Kiers & West 2015)
Diminishing returns	Social insect worker ratios (Kennedy et al. 2021), fig wasps (Wang et al. 2009)

Termitomyces

Characteristics and biogeography

Termitomyces is a genus of basidiomycete fungi of the order Agaricales, and a member of the Lyophyllaceae family. The genus is best known for its obligate mutualistic symbiosis with termites of the subfamily Macrotermitinae. *Termitomyces* species are mushroom-forming fungi, with *Termitomyces titanicus* producing the largest mushrooms known to science. Many *Termitomyces* mushrooms are edible and rich in protein, and mushrooms are highly sought after for consumption in many countries where they occur in the wild (Sangvichien & Taylor-Hawksworth 2001). In addition to its nutritional value, some species have medicinal properties (Hsieh & Ju 2018). Attempts (by humans) to cultivate *Termitomyces* mushrooms have so far been unsuccessful.

Termitomyces reproduces both sexually through basidiospores from its fruiting bodies, as well as asexually through conidiospores formed on the mycelium. Asexual reproduction facilitates spread of the fungus across substrates within a termite mound (Leuthold et al. 1989; Aanen et al. 2009), while sexual reproduction allows colonization of new host nests through spore dispersal outside the mound. Mushroom caps are equipped with a protruding tip, known as a perforatorium, to burrow through the termite mound walls. As the mycelium is often buried deep underground, mushrooms grow on a long rooting stipe (also known as a pseudorhiza) to traverse the distance to the surface. Together the rooting stipe and perforatorium form a syndrome of adaptations to subterranean growth.

Termitomyces differs from most basidiomycetes in that it appears to restrict nuclear migration during sexual reproduction (Nobre et al. 2014). Sex in basidiomycetes generally involves the fusion of post-meiotic mycelia known as homokaryons, during which both partners exchange copies of their nuclei across the fusion zone. Rather than recombining to form a diploid nucleus, the haploid nuclei from both partners remain distinct while sharing the same cell. A mycelium in this stage is known as a heterokaryon. Following nuclear migration, heterokaryon formation occurs on both sides of the fusion zone.

However, in some basidiomycetes including *Agaricus* and *Termitomyces*, the heterokaryon forms within the fusion zone itself, where both partners expel their nuclei and presumably their cytoplasm in a shared cellular mass. This indicates either an absence or a very restricted form of nuclear migration (it is noteworthy that for *Agaricus* one specific pairing did result in observable nuclear migration (Anderson et al. 1984)).

In the heterokaryon phase, *Termitomyces* also differs from most other basidiomycetes in regulating the number of nuclei per cell. Since mycelia are not strictly compartmentalized and both the cytoplasm and nuclei can move between mycelial cells, most basidiomycetes form so-called clamp connections that restrict the number of nuclei per cell to two: one copy from each parent. *Termitomyces* does not form these clamp connections and the number of nuclei per cell can vary between two and ten (De Fine Licht et al. 2005).

Termitomyces species are found mainly in sub-Saharan Africa and south-east Asia, with some species having spread overseas to Madagascar, Japan, and Indonesia. As obligate symbionts their habitat is restricted to that of their hosts, the Macrotermitinae. *Termitomyces* is thought to have originated in the rainforests of central Africa approximately 30 million years ago (Aanen & Eggleton 2005), from where it was able to invade drier habitats through its symbiotic lifestyle. Today, around 60 species have been described worldwide, although the diversity is probably much higher.

The termite-fungus symbiosis

The signature trait of *Termitomyces* is its mutualistic symbiosis with termites of the Macrotermitinae. This mutualism is obligatory and the life cycles of both fungus and termite are intricately linked through millions of years of co-evolution (Nobre et al. 2011). The mutualism has probably most aptly been described as agriculture, with the fungus acting as a crop and the termites as farmers. Insect-fungus mutualisms of this nature are also found in other insects, notably in attine ants and ambrosia beetles, as well as a few other species from other groups (Biedermann & Vega 2020). Compared to attine ants and ambrosia beetles, in which fungal domestication evolved independently several times (Biedermann & Vega 2020), the domestication of *Termitomyces* by the Macrotermitinae constitutes a single evolutionary event (Aanen et al. 2002). There has been no known reversal to a free-living state for either a *Termitomyces* species or any of its termite hosts.

Despite having an obligate dependency on each other, transmission of *Termitomyces* symbionts by most Macrotermitinae is horizontal, not vertical as might be expected for obligate mutualists (Fisher et al. 2017), and many *Termitomyces* species are associated with multiple host termite species (Aanen et al. 2002; Makonde et al. 2013; Nobre et al. 2011). Most fungus-farming termite species acquire their symbionts from the surroundings of their nest, by gathering dead plant matter inoculated with wind-dispersed basidiospores

(Johnson et al. 1981). Vertical transmission of symbionts via reproductive termites (alates) has only evolved in a small number of termite species (Korb & Aanen 2003). This vertical transmission involves storage of asexual conidiospores from the parent colony in gut compartments of either male or female (depending on the species) winged alates. These alates abandon their colony during nuptial flights and establish new colonies that are inoculated with the reposited conidiospores (Figure 2).

Termites grow fungal mycelia in special chambers of their nest on structures made of dead plant matter (so-called fungus combs). Termite workers constantly eat from both the comb and the fungus alike, and the asexual spores that form on the hyphae of *Termitomyces* survive the termite gut passage and re-inoculate the comb through their excrement. In termite species with horizontal symbiont transmission, initial fungal combs are colonized by genetically divergent *Termitomyces* homo- and heterokaryons. Due to the continuous process of asexual spore ingestion and reinoculation by termite workers (both within and between fungal combs), eventually a single fungal genotype dominates the termite fungiculture due to positive frequency-dependent selection, establishing a monoculture (Aanen et al. 2009).

While many aspects of the fungus-termite symbiosis are still unknown, there is some evidence for the principal benefits for both partners. When *Termitomyces* growing on a fungus comb is removed from a termite mound and kept in isolation, it is quickly outcompeted by other fungi of the genus *Xylaria*. The *Xylaria* fungi appear to co-exist with *Termitomyces* on the fungus comb, but either remain dormant or are otherwise kept from outcompeting *Termitomyces* under natural conditions (Visser et al. 2011, 2009). This indicates *Termitomyces* is dependent on the termites for outcompeting other fungi. In addition, *Termitomyces* is abundant in habitats that are dry and probably unsuitable for its growth outside the shielded environment of a termite mound. Its association with termites has allowed *Termitomyces* to colonize areas that would otherwise have been inaccessible.

The termites benefit from the fungal ability to degrade lignocellulose. Most termite genera possess gut protists that enable them to digest lignocellulose in plant cells, but the Macrotermitinae lack these gut microbes. Comparative enzymatic studies of *Termitomyces*, termite and termite gut bacteria show *Termitomyces* can act as an external digestive tract, breaking down lignin in the plant matter making up a fungus comb, leaving sugars for the termites (Poulsen et al. 2014). A recent study shows that although *Termitomyces* lacks a full complement of enzymes for lignin degradation as found in most white-rot fungi, it may use a specialized Fenton-based pathway of oxidative degradation to compensate (Schalk et al. 2021). As Fenton reaction-based lignin digestion is commonly associated with brown-rot fungi, this would imply that *Termitomyces* combines digestive strategies from both white- and brown-rot fungi to break down lignin.

Ecologically, fungus-farming termites are major contributors to plant decomposition in their habitats, significantly impacting soil composition and fertility in their environment. This also makes them important players in (global) carbon cycling as almost all carbon from their food sources is expelled as carbon-dioxide and methane gas, contributing to the greenhouse effect (Jones 1990). Conversely, termite mound-building may provide a buffer for local vegetation against drought and desertification (Bonachela et al. 2015). In addition, termites enrich soil with excess nutrients such as nitrogen and phosphorus, indirectly increasing the biodiversity of their ecosystems (Pennisi 2015). Many termite species associated with *Termitomyces* are considered pests and cause significant damage to agriculture and structures. Given that these termite species depend utterly on their fungal symbionts for survival, taking countermeasures against the fungus can complement countermeasures taken against the termites themselves.

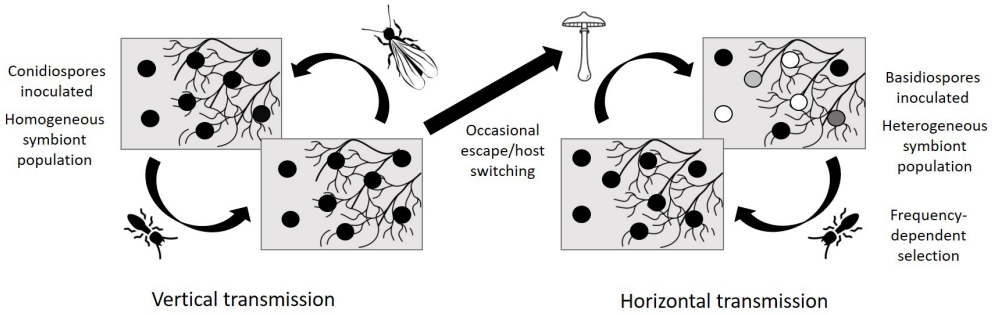


Figure 2: Transmission of *Termitomyces*

The transmission of *Termitomyces* to new termite colonies occurs either vertically (left) or horizontally (right) depending on the host termite species. Horizontal transmission is more common and involves wind-dispersed basidiospores from *Termitomyces* mushrooms that usually break out through the walls of their host termite mound. Foraging termite workers of newly established colonies inoculate basidiospores from their environment on fresh fungus combs. The spores germinate and mate with a compatible partner, forming heterokaryons that spread through a mixture of fusion, hyphal growth and conidiospore formation. Termite workers continuously consume fungus and comb alike, reinoculating the comb with the conidiospores that survive the gut passage. Over time, this grazing and reseeding results in a single heterokaryon strain dominating the fungal population by positive frequency-dependent selection.

Vertical transmission of *Termitomyces* evolved in a few species of the Macrotermitinae. It bypasses the sexual reproduction of *Termitomyces* by transporting conidiospores from the parent colony to new colonies through the reproductive termite males or females. This maintains the genetic homogeneity of the fungal crop by restricting it to asexual reproduction, whereas the horizontal transmission mode allows intermittent sex each time a new termite colony is established. However,

vertically transmitted *Termitomyces* species are probably not completely restricted to asexual reproduction, as they may be associated with multiple host species, only some of which are adapted to vertical symbiont transmission (Nobre et al. 2014)

Mitochondrial DNA

Mitochondria were once free-living bacteria, likely belonging to the Rickettsiae (Emelyanov 2001), that became endosymbionts of an ancestor of modern eukaryotes (Sagan 1967). The separate origin of mitochondria from the eukaryote cell nucleus is exemplified by a residual mitochondrial genome (mtDNA), often encoded with a different genetic code than the host nuclear DNA. This genome is reduced to various degrees across the eukaryotes, with most DNA transferred to the host nucleus or simply lost. A core set of genes involved in the respiratory metabolism remains conserved in virtually all mitochondrial genomes, including prerequisite tRNA and rRNA genes. In humans and other bilaterian animals, the mtDNA consists of not much other than this core set of mitochondrial genes, with almost no intergenic space remaining and no known introns (Figure 3) (Lavrov & Pett 2016). In other eukaryotes, there is much more diversity in mtDNA content (Lavrov & Pett 2016; Burger et al. 2003; Bullerwell & Gray 2004).

Fungal mitochondrial DNA varies in size between around 20-300kb (Liu et al. 2020; Freel et al. 2015), generally encoding 14 protein-coding genes, two ribosomal subunits and 20-odd tRNA genes (Figure 3). In addition, many fungal mitogenomes contain mobile introns (homing endonucleases or reverse transcriptases) (Haugen et al. 2005), repetitive elements and sequences of retroviral plasmid origin (Cahan & Kennell 2005), which together account for much of the size variation observed for fungal mtDNA. Fungal mitochondrial genomes also vary in structure, although most seem to conform to the 'standard' head-tail bound concatameric genomes (Bendich 1996), linear genomes capped with inverted repeats have also been reported (Forget et al. 2002). This variation in mtDNA content and structure demands an equal diversity in genome regulation and replication (Burger et al. 2003), making fungal mtDNA a resourceful area to study fundamental aspects of (mitochondrial) genome biology.

Mitochondrial inheritance during sexual reproduction is non-Mendelian, with usually only one parent passing on its mtDNA to the offspring. In anisogamous species, where gametes are differentiated by sex, mitochondria are virtually always inherited from the maternal side (i.e., the sex that produces the largest gamete). In isogamous species uniparental inheritance can also occur, controlled for example by mating type-determining loci or stochastic segregation and turnover of mtDNA in heteroplasmic cells (Barr et al. 2005). In fungi, uniparental inheritance is observed in many anisogamous and isogamous species,

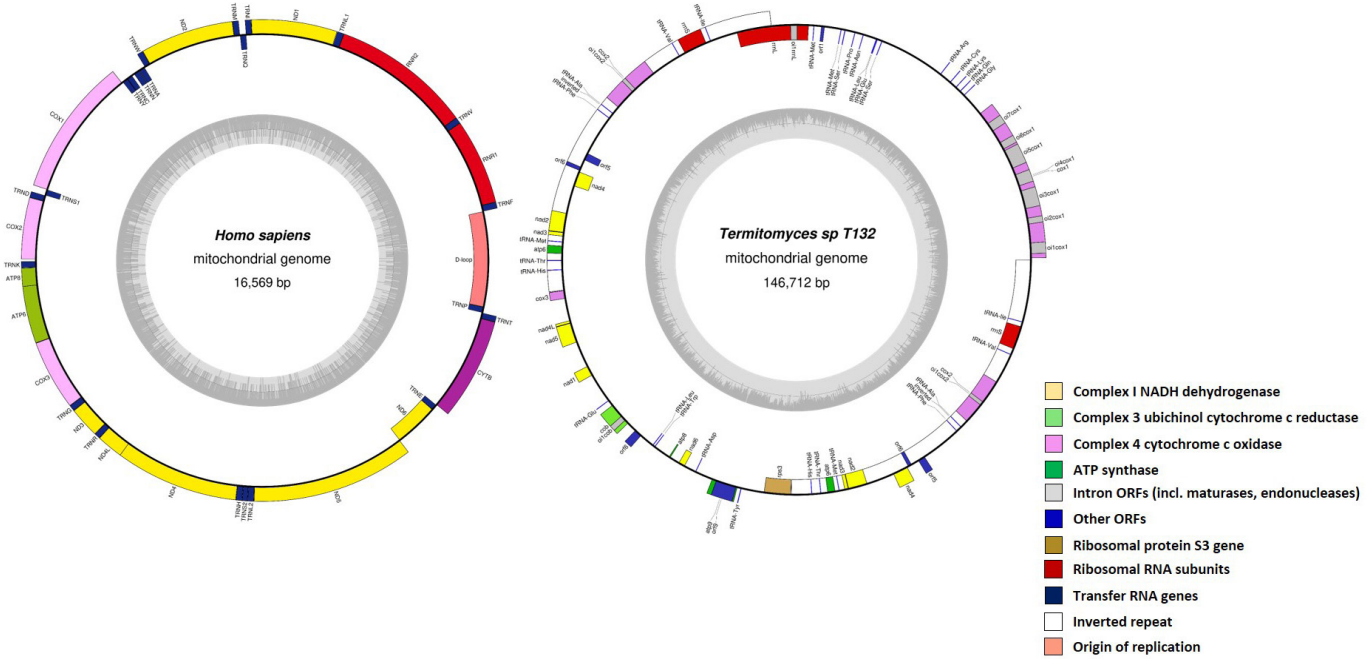


Figure 3: Mitochondrial genome map of *Homo sapiens* and *Termitomyces* sp. T132
The mitochondrial genomes of bilaterian animals like *Homo sapiens* are typically small (around 16 kb), condensed genomes with almost no non-coding DNA. However, they preserve the core set of mitochondrial genes found in almost all eukaryote mtDNA: 14 protein-coding genes, 2 ribosomal RNA subunits and around two dozen tRNA genes. By contrast, fungal mtDNA like that of Termitomyces sp. T132 contains much more non-coding DNA such as introns (including type I and type II mobile introns), large intergenic regions and repetitive DNA. As a result, depending on the amount of non-coding DNA in the genome the mtDNA of fungal species varies in size from around 20kb to 300kb.

but biparental inheritance is also known to occur in a number of species (Barr et al. 2005; Wang et al. 2017). The preponderance of uniparental transmission of mtDNA agrees with the general pattern of hosts restricting the genetic variation of their (obligate) symbionts (Frank 1996; Hoekstra 2000). Limiting genetic variation of symbiont populations reduces the risk of selfish variants arising that could outcompete other, more altruistic symbionts (Hastings 1992). Biparental inheritance can facilitate recombination of genetically divergent mitochondrial genomes, and could potentially result in selection for selfish mtDNA variants that are a detriment to the host cell (Christie et al. 2015). Furthermore, there is some evidence for a direct detrimental effect of heteroplasmy to the host (Sharpley et al. 2012).

In fungi, mobile genetic elements such as homing endonucleases are common residents of mtDNA (Hausner 2003). Homing endonucleases use self-splicing pathways to recombine separately from the rest of the mitogenome (Haugen et al. 2005; Delahodde et al. 1989). Studies of fungal mtDNA recombination should therefore consider homing activity as a potential factor. In yeast, observations of mtDNA recombination were found to be cases of homing endonuclease activity (Xu et al. 2009; Wu & Hao 2019).

Mitochondrial plasmids

In fungi and plants, mitochondria are frequently host to various autonomous plasmids. These plasmids are thought to be as ancient as the mitochondrial endosymbiosis itself, possibly deriving from bacteriophages (Shutt & Gray 2006). Evidence for this comes from genetic homology between these plasmids and T-odd bacteriophages. In addition, the nuclear gene regulating mitochondrial replication in eukaryotes also shares homology with the RNA polymerase carried by these plasmids and T-odd bacteriophages, and homologs of this gene are also found in alpha-proteobacterial genomes (Shutt & Gray 2006). This suggests that the origin of the obligate mutualism between mitochondrial and nuclear DNA is evolutionarily connected to these plasmids and phages.

Mitochondrial plasmids can be divided into circular and linear types, of which the linear type is believed to be more common (Griffiths 1995). Both types can insert themselves in their host mtDNA, and plasmid-derived DNA fragments are frequently found in fungal mtDNA sequences (Ferandon et al. 2008; Férandon et al. 2013; Himmelstrand et al. 2014; Hausner 2003). Transfer of mtDNA to plasmids has been reported for circular plasmids (Akins et al. 1989; Mohr et al. 2000) and a linear plasmid found in maize (Leon et al. 1989), but not in fungal linear plasmids. In the reported cases, transfer usually involved the duplication of a complete tRNA gene to the plasmid. In the case of the maize plasmid, the presence of a mitochondrial tRNA gene on a plasmid coincided with a loss of that tRNA gene from the plant's mtDNA. This potentially creates a form of gene addiction (Kobayashi

2004) of the plant to the plasmid, assuming the tRNA is essential and only available from the plasmid.

Thesis outline

The origin of the fungus-termite mutualism

The specialized lifestyle of *Termitomyces* raises many questions about its evolution and, more fundamentally, the evolution of long-term symbioses in general. We do not know how *Termitomyces* became domesticated by the Macrotermitinae. Domestications are relatively rare mutualistic interactions that have proven to be highly successful ecological strategies allowing for widespread dispersal and colonization of otherwise unsuitable habitats, for example in humans and the Macrotermitinae, and their respective symbiont crops (Milla et al. 2015; Zeder 2016; Aanen & Eggleton 2005). There are many unanswered questions regarding the origin and impact of domestication events in human evolutionary history (Larson et al. 2014), let alone those in other organisms. Studying the evolution of domestication across different host and symbiont species can reveal common characteristics that may be fundamental to their success.

By comparing the ecology and lifestyle of species closely related to *Termitomyces*, we can attempt to reconstruct the ancestral state of their common ancestors and trace back the steps towards the fungus-termite mutualism. For this we require a phylogeny that includes as many of the closest known relatives to *Termitomyces* as possible, and a representative set of *Termitomyces* species themselves, as well as ecological and morphological data relevant to the ancestral condition for all of these species. In Chapter 3, I utilize a genomic data set of more than 1100 loci for 25 *Termitomyces* species and 21 close relatives within the Lyophyllaceae to reconstruct their phylogeny. I complement this phylogeny with ecological, morphological and enzymatic trait data to estimate the lifestyles of *Termitomyces'* relatives and the ancestral conditions that preceded and perhaps facilitated the domestication by termites.

Mitochondrial evolution in Termitomyces

Termitomyces shows unusual heterokaryon formation during sexual reproduction consistent with limited nuclear migration. We do not know the consequences of this mating process for mitochondrial inheritance, and whether it enables recombination of mitochondrial DNA between different parental mitochondria in the fusion zone. Biparental inheritance (BPI) and recombination of mtDNA present a potential source for genomic conflicts between nuclear and mitochondrial genomes (Christie et al. 2015). I am interested in the possibility of BPI and recombinant mtDNA in *Termitomyces* and the consequences for mitochondrial genome evolution in this genus. In Chapter 5, I

combine mitochondrial sequence data from experimental matings in the lab and from wild strains of a *Termitomyces* species (*T. cryptogamus*). I use this combined data set to test the hypothesis that limited nuclear migration in *T. cryptogamus* results in biparental inheritance and recombination of mtDNA.

As a foundation for studying mitochondrial evolution and inheritance in *Termitomyces*, I assembled complete mitochondrial genomes for several species of the genus and its closest relatives, comparing their structure and reconstructing a mitochondrial phylogeny based on fourteen core protein-coding genes. Since no complete mtDNA sequences from the Lyophyllaceae had previously been published, several unanticipated characteristics (namely a set of large, inverted repeats and a high frequency of potential G-quadruplex motifs) of some of the mtDNA sequences raised further questions on their evolutionary history and possible function. I describe the mitochondrial genome assemblies in Chapter 2, in addition to comparing several of their features to other fungal mitogenomes and also human mtDNA.

During the assembly of *Termitomyces* mitochondrial genomes I also discovered the sequences of numerous linear mitochondrial plasmids in my samples. These plasmids are frequently found in plant and fungal mtDNA and have several intriguing capabilities, including horizontal transfer across species barriers, insertion into mitochondrial DNA and, in some specific cases, inducing senescence in their fungal host (Griffiths 1995). There have also been rare cases reported of mtDNA transfer, in particular tRNA genes, to certain plasmids (Akins et al. 1989; Kempken 1989; Leon et al. 1989; Mohr et al. 2000). In Chapter 4, I screened all whole-genome assemblies generated in Chapter 3 for plasmid sequences and used the resulting data set to analyze the evolutionary relationships among them and those previously reported in other fungi, and describe several cases of mtDNA transfer to plasmids.

References

Aanen DK et al. 2009. High symbiont relatedness stabilizes mutualistic cooperation in fungus-growing termites. Science. 326:1103–1106. doi: 10.1126/science.1173462.

Aanen DK et al. 2002. The evolution of fungus-growing termites and their mutualistic fungal symbionts. Proc. Natl. Acad. Sci. U. S. A. 99:14887–14892. doi: 10.1073/pnas.222313099.

Aanen DK, Eggleton P. 2005. Fungus-growing termites originated in African rain forest. Curr. Biol. 15:851–855. doi: 10.1016/j.cub.2005.03.043.

Akins RA, Kelley RL, Lambowitz AM. 1989. Characterization of mutant mitochondrial plasmids of Neurospora spp. that have incorporated tRNAs by reverse transcription. Mol. Cell. Biol. 9:678–691. doi: 10.1128/mcb.9.2.678.

Anderson JB, Petsche DM, Herr FB, Horgen PA. 1984. Breeding relationships among several species of Agaricus . Can. J. Bot. 62:1884–1889. doi: 10.1139/b84-257.

Axelrod R, Hamilton WD. 1981. The evolution of cooperation. Science (80-.). 211:1390–1396. doi: 10.1126/science.7466396.

Barr CM, Neiman M, Taylor DR. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytol. 168:39–50. doi: 10.1111/j.1469-8137.2005.01492.x.

Bendich AJ. 1996. Structural analysis of mitochondrial DNA molecules from fungi and plants using moving pictures and pulsed-field gel electrophoresis. J. Mol. Biol. 255:564–588. doi: 10.1006/jmbi.1996.0048.

Biedermann PHW, Vega FE. 2020. Ecology and Evolution of Insect–Fungus Mutualisms. Annu. Rev. Entomol. 65:431–455. doi: 10.1146/annurev-ento-011019-024910.

Birky CW. 2001. The Inheritance of Genes in Mitochondria and Chloroplasts: Laws, Mechanisms, and Models. Annu. Rev. Genet. 35:125–148. doi: 10.1146/annurev.genet.35.102401.090231.

Bonachela JA et al. 2015. Termite mounds can increase the robustness of dryland ecosystems to climatic change. Science (80-.). 347:651–655. doi: 10.1126/science.1261487.

Bullerwell CE, Gray MW. 2004. Evolution of the mitochondrial genome: protist connections to animals, fungi and plants. Curr. Opin. Microbiol. 7:528–534. doi: https://doi.org/10.1016/j.mib.2004.08.008.

Burger G, Gray MW, Lang BF. 2003. Mitochondrial genomes: Anything goes. Trends Genet. 19:709–716. doi: 10.1016/j.tig.2003.10.012.

Cahan P, Kennell JC. 2005. Identification and distribution of sequences having similarity to mitochondrial plasmids in mitochondrial genomes of filamentous fungi. Mol. Genet. Genomics. 273:462–473. doi: 10.1007/s00438-005-1133-x.

Cassar S, Blackwell M. 1996. Convergent origins of ambrosia fungi. Mycologia. 88:596–601.

Christie JR, Schaerf TM, Beekman M. 2015. Selection against Heteroplasmy Explains the Evolution of Uniparental Inheritance of Mitochondria. PLoS Genet. 11:1–30. doi: 10.1371/journal.pgen.1005112.

Darwin C. 1859. *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. London : John Murray, 1859 https://search.library.wisc.edu/catalog/9934839413602122.

Delahodde A et al. 1989. Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. Cell. 56:431–441. doi: 10.1016/0092-8674(89)90246-8.

Emelyanov V V. 2001. Evolutionary relationship of Rickettsiae and mitochondria. FEBS Lett. 501:11–18. doi: https://doi.org/10.1016/S0014-5793(01)02618-7.

Enticknap JJ, Kelly M, Peraud O, Hill RT. 2006. Characterization of a Culturable Alphaproteobacterial Symbiont Common to Many Marine Sponges and Evidence for Vertical Transmission via Sponge Larvae. Appl. Environ. Microbiol. 72:3724–3732. doi: 10.1128/AEM.72.5.3724-3732.2006.

Ferandon C, Chatel SEK, Castandet B, Castroviejo M, Barroso G. 2008. The Agrocybe aegerita mitochondrial genome contains two inverted repeats of the nad4 gene arisen by duplication on both sides of a linear plasmid integration site. Fungal Genet. Biol. 45:292–301. doi: 10.1016/j.fgb.2007.10.012.

Férandon C, Xu J, Barroso G. 2013. The 135 kbp mitochondrial genome of Agaricus bisporus is the largest known eukaryotic reservoir of group I introns and plasmid-related sequences. Fungal Genet. Biol. 55:85–91. doi: 10.1016/j.fgb.2013.01.009.

De Fine Licht HH, Andersen A, Aanen DK. 2005. Termitomyces sp. associated with the termite Macrotermes natalensis has a heterothallic mating system and multinucleate cells. Mycol. Res. 109:314–318. doi: 10.1017/S0953756204001844.

Fisher RA. 1930. *The genetical theory of natural selection.* Clarendon Press: Oxford, England doi: 10.5962/bhl.title.27468.

Fisher RM, Henry LM, Cornwallis CK, Kiers ET, West SA. 2017. The evolution of host-symbiont dependence. Nat. Commun. 8:15973. doi: 10.1038/ncomms15973.

Forget L, Ustinova J, Wang Z, Huss VAR, Franz Lang B. 2002. Hyaloraphidium curvatum: A Linear Mitochondrial Genome, tRNA Editing, and an Evolutionary Link to Lower Fungi. Mol. Biol. Evol. 19:310–319. doi: 10.1093/oxfordjournals.molbev.a004084.

Frank SA. 1996. Host Control of Symbiont Transmission: The Separation of Symbionts Into Germ and Soma. Am. Nat. 148:1113–1124. doi: 10.1086/285974.

Freel KC, Friedrich A, Schacherer J. 2015. Mitochondrial genome evolution in yeasts: An all-encompassing view. FEMS Yeast Res. 15:1–9. doi: 10.1093/femsyr/fov023.

Griffiths AJF. 1995. Natural plasmids of filamentous fungi. Microbiol. Rev. 59:673–685. doi: 10.1128/mmbr.59.4.673-685.1995.

Haldane JBS. 1932. *The causes of evolution.* Macmillan: Oxford, England.

Hamilton WD. 1964. The genetical evolution of social behaviour. I. J. Theor. Biol. 7:1–16. doi: https://doi.org/10.1016/0022-5193(64)90038-4.

Hardin G. 1968. The Tragedy of the Commons. Science (80-.).

Hastings IM. 1992. Population genetic aspects of deleterious cytoplasmic genomes and their effect on the evolution of sexual reproduction. Genet. Res. 59:215–225. doi: 10.1017/S0016672300030500.

Haugen P, Simon DM, Bhattacharya D. 2005. The natural history of group I introns. Trends Genet. 21:111–119. doi: https://doi.org/10.1016/j.tig.2004.12.007.

Hausner G. 2003. Fungal Mitochondrial Genomes. Fungal Genomics. 3:101.

Himmelstrand K, Olson Å, Brandström Durling M, Karlsson M, Stenlid J. 2014. Intronic and plasmid-derived regions contribute to the large mitochondrial genome sizes of Agaricomycetes. Curr. Genet. 60:303–313. doi: 10.1007/s00294-014-0436-z.

Hoekstra RF. 2000. Evolutionary origin and consequences of uniparental mitochondrial inheritance. Hum. Reprod. 15:102–111. doi: 10.1093/humrep/15.suppl_2.102.

Hsieh H-M, Ju Y-M. 2018. Medicinal components in Termitomyces mushrooms. Appl. Microbiol. Biotechnol. 102:4987–4994. doi: 10.1007/s00253-018-8991-8.

Johnson RA, Thomas RJ, Wood TG, Swift MJ. 1981. The inoculation of the fungus comb in newly founded colonies of some species of the Macrotermitinae (Isoptera) from Nigeria. J. Nat. Hist. 15:751–756. doi: 10.1080/00222938100770541.

Jones JA. 1990. Termites, Soil Fertility and Carbon Cycling in Dry Tropical Africa: A Hypothesis. J. Trop. Ecol. 6:291–305. doi: 10.1017/S0266467400004533.

Kempken F. 1989. A tRNA-like structure from a linear plasmid of the fungus Ascobolus immersus. Nucleic Acids Res. 17:6740. doi: 10.1093/nar/17.16.6740.

Kennedy P et al. 2021. Diminishing returns drive altruists to help extended family. Nat. Ecol. Evol. 5:468–479. doi: 10.1038/s41559-020-01382-z.

Kiers ET, West SA. 2015. Evolving new organisms via symbiosis. Science (80-.). 348:392–394. doi: 10.1126/science.aaa9605.

Kim I, Rodriguez-Enriquez S, Lemasters JJ. 2007. Selective degradation of mitochondria by mitophagy. Arch. Biochem. Biophys. 462:245–253. doi: https://doi.org/10.1016/j.abb.2007.03.034.

Kobayashi I. 2004. Genetic Addiction: a Principle of Gene Symbiosis in a Genome. Plasmid Biol. 105–144. doi: https://doi.org/10.1128/9781555817732.ch6.

Koga R, Meng X-Y, Tsuchida T, Fukatsu T. 2012. Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte{\textendash}embryo interface. Proc. Natl. Acad. Sci. 109:E1230--E1237. doi: 10.1073/pnas.1119212109.

Korb J, Aanen DK. 2003. The evolution of uniparental transmission of fungal symbionts in fungus-growing termites (Macrotermitinae). Behav. Ecol. Sociobiol. 53:65–71. doi: 10.1007/s00265-002-0559-y.

Larson G et al. 2014. Current perspectives and the future of domestication studies. Proc. Natl. Acad. Sci. 111:6139–6146. doi: 10.1073/pnas.1323964111.

Lavrov D V., Pett W. 2016. Animal mitochondrial DNA as we do not know it: Mt-Genome Organization and Evolution in Nonbilaterian Lineages. Genome Biol. Evol. 8:2896–2913. doi: 10.1093/gbe/evw195.

Leigh Jr EG. 2010. The evolution of mutualism. J. Evol. Biol. 23:2507–2528. doi: https://doi.org/10.1111/j.1420-9101.2010.02114.x.

Leon P, Walbot V, Bedinger P. 1989. Molecular analysis of the linear 2.3 kb plasmaid of maize mitochondria: apparent capture of tRNA genes. Nucleic Acids Res. 17:4089–4099.

Leuthold RH, Badertscher S, Imboden H. 1989. The inoculation of newly formed fungus comb withTermitomyces inMacrotermes colonies (Isoptera, Macrotermitinae). Insectes Soc. 36:328–338. doi: 10.1007/BF02224884.

Liu W et al. 2020. The mitochondrial genome of Morchella importuna (272.2 kb) is the largest among fungi and contains numerous introns, mitochondrial non-conserved open reading frames and repetitive sequences. Int. J. Biol. Macromol. 143:373–381. doi: https://doi.org/10.1016/j.ijbiomac.2019.12.056.

Makonde HM et al. 2013. Diversity of Termitomyces Associated with Fungus-Farming Termites Assessed by Cultural and Culture-Independent Methods. PLoS One. 8. doi: 10.1371/journal.pone.0056464.

Maynard Smith J, Szathmáry E. 1997. The major transitions in evolution. http://site.ebrary.com/id/10464134.

Milla R, Osborne CP, Turcotte MM, Violle C. 2015. Plant domestication through an ecological lens. Trends Ecol. Evol. 30:463–469. doi: https://doi.org/10.1016/j.tree.2015.06.006.

Mohr S, Wanner LA, Bertrand H, Lambowitz AM. 2000. Characterization of an unusual tRNA-like sequence found inserted in a Neurospora retroplasmid. Nucleic Acids Res. 28:1514–1524. doi: 10.1093/nar/28.7.1514.

Mueller UG, Schultz TR, Currie CR, Malloch D. 2001. The Origin of the Attine Ant-Fungus Mutualism. Q. Rev. Biol. 76:169–197. doi: 10.1086/393867.

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. 20:2619–2627. doi: 10.1111/j.1365-294X.2011.05090.x.

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. 14:1–12. doi: 10.1186/1471-2148-14-121.

Nowak MA. 2006. Five Rules for the Evolution of Cooperation. Science (80-.). 314:1560–1563. doi: 10.1126/science.1133755.

Pennisi E. 2015. Africa{\textquoteright}s soil engineers: Termites. Science (80-.). 347:596–597. doi: 10.1126/science.347.6222.596.

Poulsen M et al. 2014. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. Proc. Natl. Acad. Sci. U. S. A. 111:14500–14505. doi: 10.1073/pnas.1319718111.

Rankin DJ, Bargum K, Kokko H. 2007. The tragedy of the commons in evolutionary biology. Trends Ecol. Evol. 22:643–651. doi: 10.1016/j.tree.2007.07.009.

Sagan L. 1967. On the origin of mitosing cells. J. Theor. Biol. 14:225-IN6. doi: https://doi.org/10.1016/0022-5193(67)90079-3.

Sangvichien E, Taylor-Hawksworth P. 2001. Termitomyces mushrooms: A tropical delicacy. Mycologist. 15:31–33. doi: 10.1016/S0269-915X(01)80058-6.

Schalk F et al. 2021. The termite fungal cultivar Termitomyces combines diverse enzymes and oxidative reactions for plant biomass conversion. bioRxiv. doi: 10.1101/2021.01.13.426627.

Sharpley MS et al. 2012. Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and cognition. Cell. 151:333–343. doi: 10.1016/j.cell.2012.09.004.

Shutt TE, Gray MW. 2006. Bacteriophage origins of mitochondrial replication and transcription proteins. Trends Genet. 22:90–95. doi: 10.1016/j.tig.2005.11.007.

Smith JM. 1964. Group Selection and Kin Selection. Nature. 201:1145–1147. doi: 10.1038/2011145a0.

Smith JM, Szathmary E. 1997. *The major transitions in evolution*. Oxford University Press.

Trivers RL. 1971. The Evolution of Reciprocal Altruism. Q. Rev. Biol. 46:35–57. doi: 10.1086/406755.

Visick KL, Foster J, Doino J, McFall-Ngai M, Ruby EG. 2000. Vibrio fischeri lux Genes Play an Important Role in Colonization and Development of the Host Light Organ. J. Bacteriol. 182:4578–4586. doi: 10.1128/JB.182.16.4578-4586.2000.

Visser AA et al. 2009. Levels of specificity of Xylaria species associated with fungus-growing termites: a phylogenetic approach. Mol. Ecol. 18:553–567. doi: https://doi.org/10.1111/j.1365-294X.2008.04036.x.

Visser AA, Kooij PW, Debets AJM, Kuyper TW, Aanen DK. 2011. Pseudoxylaria as stowaway of the fungus-growing termite nest: Interaction asymmetry between Pseudoxylaria, Termitomyces and free-living relatives. Fungal Ecol. 4:322–332. doi: 10.1016/j.funeco.2011.05.003.

Wang P et al. 2017. Frequent heteroplasmy and recombination in the mitochondrial genomes of the basidiomycete mushroom Thelephora ganbajun. Sci. Rep. 7:1–14. doi: 10.1038/s41598-017-01823-z.

Wang R-W et al. 2009. Interference Competition and High Temperatures Reduce the Virulence of Fig Wasps and Stabilize a Fig-Wasp Mutualism. PLoS One. 4:e7802. https://doi.org/10.1371/journal.pone.0007802.

Wu B, Hao W. 2019. Mitochondrial-encoded endonucleases drive recombination of protein-coding genes in yeast. Environ. Microbiol. 21:4233–4240. doi: https://doi.org/10.1111/1462-2920.14783.

Xu J, Yan Z, Guo H. 2009. Divergence, hybridization, and recombination in the mitochondrial genome of the human pathogenic yeast Cryptococcus gattii. Mol. Ecol. 18:2628–2642. doi: 10.1111/j.1365-294X.2009.04227.x.

Zeder MA. 2016. Domestication as a model system for niche construction theory. Evol. Ecol. 30:325–348. doi: 10.1007/s10682-015-9801-8.

CHAPTER 2

2

Enrichment of G4DNA and a large inverted repeat coincide in the mitochondrial genomes of *Termitomyces*

Mathijs Nieuwenhuis[1*], Lennart JJ van de Peppel[1], Freek T Bakker[2], Bas J Zwaan[1], Duur K Aanen[1]

[1] *Laboratory of Genetics, Wageningen University & Research, Droevendaalsesteeg 1, Radix West, Building 107 6708 PB Wageningen, the Netherlands*

[2] *Biosystematics Group, Wageningen University & Research, Droevendaalsesteeg 1, Radix West, Building 107 6708 PB Wageningen, the Netherlands*

* *Author for Correspondence; E-mail: mathijs.nieuwenhuis@wur.nl*

Published in Genome Biology & Evolution, Volume 11, Issue 7, July 2019, Pages 1857–1869

Summary

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 eleven 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 characterised by a large inverted repeat 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

Data Deposition: MG783568, MH725791, MH725792, MH725793, MH725794, MH725795, MH725796, MH725797, MH725798, MH725799, MH725800, MH743217

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 & Xu 2012; Smith & 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 fertilised 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, non-recombining 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 date no mtDNA sequence for *Termitomyces* or any closely related fungus has been published. In this paper, 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 into 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 (Johnson et al. 2009; Koole et al. 2014; Verma et al. 2009; Schaffitzel et al. 2001). 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; Cui et al. 2016; Verma et al. 2009; Bharti et al. 2014; Huang et al. 2015; Rawal et al. 2006). In addition, the telomeres of many eukaryotes are comprised of 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 (Valton et al. 2014; Paeschke et al. 2012). G4DNA forms more frequently in cancer cells (Biffi et al. 2014), and G4DNA motifs are prevalent in promotors of oncogenes (Eddy & 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 & Maizels 2006; Rawal et al. 2006; Yadav et al. 2008; Du et al. 2008; Garg et al. 2016; Capra et al. 2010). 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.* (Capra et al. 2010) analysed the G4DNA motif content of the nuclear and mitochondrial genome of *Saccharomyces cerevisiae,* and found an approximately tenfold increase of G4DNA motif content in mtDNA compared to 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 (Turmel et al. 1999; Palmer 1985). 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 non-duplicated regions (Aldrich et al. 1985; Palmer 1983). Inverted repeats 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 & Thompson 1982; Maier et al. 1995), but in other studies this effect was not observed (Blazier et al. 2016). HR-driven repair of double-stranded breaks (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 centred 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 paper, we present new mitochondrial genome sequences of twelve 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.

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 *Tephrocybe 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 *Asterophora 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 litre demi water: 20g malt extract, 2g yeast extract, 15g agar). The *Termitomyces* isolates were grown at 25°C in the dark. *Tricholomella constricta*, *Asterophora parasitica*, *Tephrocybe rancida* and *Myochromella 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). Library preparation and whole-genome sequencing was performed by Novogene (Hong Kong) using the Illumina Hiseq 2500 platform. Generated reads were 150bp long and the insert size was 500bp. Because of the high coverage of

mtDNA, we generally used a subsample of 1-2 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 inverted repeat as follows: 1) contigs covering the IR were identified by a two-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 artefacts due to long (10bp+) 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 (non-conserved) 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) was 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 & 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 inverted repeat 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). 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). 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 (c*ox1, cox2, cox3, cob* and *atp6*). We used RevTrans (Wernersson & 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 5000 bootstraps and otherwise default settings (Nguyen et al. 2015). For the fungal phylogeny, we included a topological constraint tree (supplementary data 6) 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 Cipres web server (Miller et al. 2010) for the Lyophyllaceae phylogeny with the following parameter settings: 50 million generations, sample frequency 5000, 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 paper 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 non-silent 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 more than 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 (X. Xia and Z. 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 twelve 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 et al., unpublished data), *Blastosporella zonata, Tephrocybe rancida, Myochromella boudieri, Tricholomella constricta,* and *Asterophora 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 inverted repeat region and therefore duplication of all genes contained within, though the gene content of the IR varied from species to species. All species except *A. parasitica* had introns in one or more mitochondrial genes. In particular, *cox1* often featured several introns, most of which harbour LAGLIDADG or GIY-YIG endonucleases, a phenomenon commonly found in basidiomycete fungi (Lang et al. 2007).

Mitochondrial inverted repeat

We detected the mitochondrial inverted repeat in our assemblies by a two-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 pairings between the ends of the assembled sequence and the borders of the duplicated region. To confirm the presence of the inverted repeat, 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). 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 23s 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.

Phylogenetic analysis/Mauve

Our phylogenetic analysis (Figure 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 *A. 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*. *M. 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 (Nobre et al. 2011; Aanen et al. 2002). 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.

Table 1: Overview of mitochondrial genomes of twelve Lyophyllaceae species assembled and annotated for this study.

					Introns									
Strain	**GenBank ID**	**mtDNA size (bp)**	**IR size (bp)**	**GC content (%)**	**atp9**	**cob**	**cox1**	**cox2**	**nad1**	**nad2**	**nad4**	**nad5**	**rns**	**rnl**
Asterophora parasitica	MH725791	43328	0	31.7	0	0	0	0	0	0	0	0	0	2
Tricholomella constricta	MH725800	65087	0	25.3	0	1	0	0	0	0	0	0	0	2
Myochromella boudieri	MH725793	99774	0	27.9	0	3	9	2	2	0	0	1	0	4
Blastosporella zonata	MH725792	200401	56326	34.8	1	6	6	3	0	0	1	2	2	4
Tephrocybe rancida	MH725794	126794	33754	38.3	0	2	3	2	0	0	0	0	0	1
Termitomyces sp. Mi166	MH725795	239317	71805	37.8	1	4	15	1	1	1	0	0	2	8
Termitomyces sp. T159	MH725799	157156	41609	39.3	1	2	4	1	0	0	0	0	1	3
Termitomyces sp. T32	MH725797	131333	35619	40.9	0	0	5	2	0	0	0	0	0	1
Termitomyces sp. T13	MH725796	155430	44175	34.5	0	3	6	2	0	0	0	0	2	2
Termitomyces sp. DKA19	MH743217	105724	4463	31	1	2	8	2	0	0	0	0	0	1
Termitomyces sp. T123	MH725798	124711	28228	33.1	0	1	6	2	0	1	0	1	0	3
Termitomyces sp. T132	MG783568	146712	37211	31.3	1	1	8	3	0	0	0	0	0	2

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, e.g. maranhar/kalilo in *Neurospora* (Chan et al. 1991; Court & 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 inverted repeats (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 (~two-fold higher) coverage compared to 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 (<<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.

Table 2: Genes in IR for each species; asterisks indicate genes partially overlapping with IRs. Also shown are flanking genes for each set of border regions (IR-SC1, IR-SC2).

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

We compared silent substitution rate estimates of genes of the twelve 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). Ka/Ks 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 harboured the replication origins as was found for *Candida albicans* (Gerhold et al. 2010) we analysed the cumulative GC skew profile for each mitochondrial genome with an IR (supplementary data 5). 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.

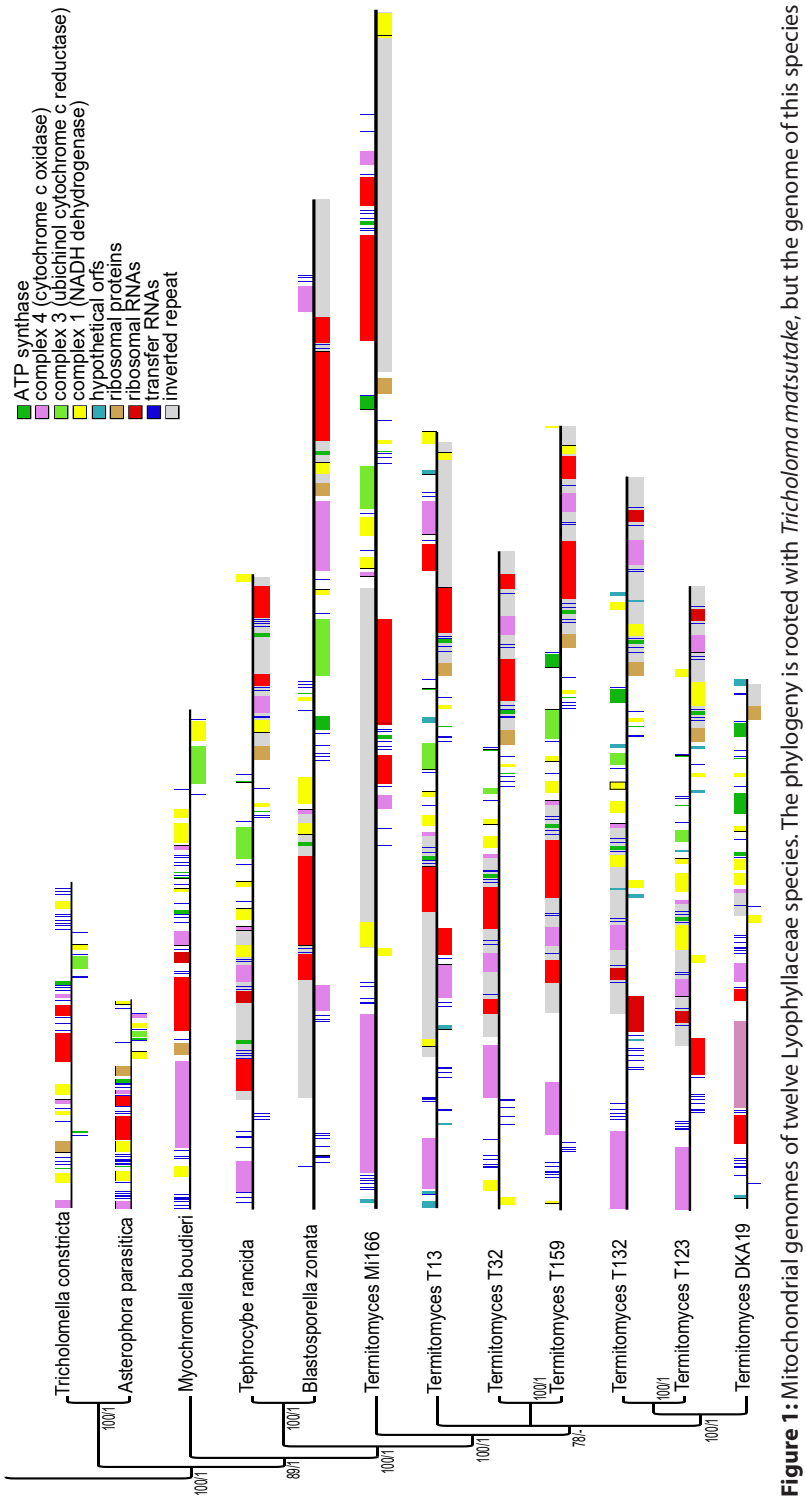
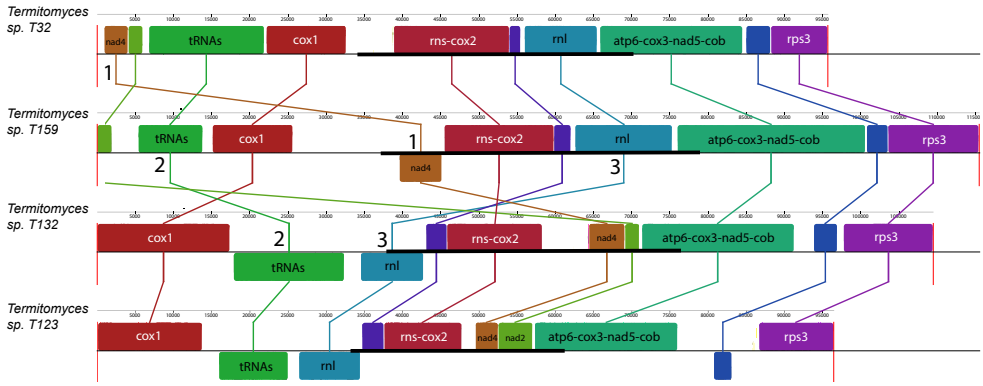


Figure 1: Mitochondrial genomes of twelve Lyophyllaceae species. The phylogeny is rooted with *Tricholoma matsutake*, but the genome of this species is not included in the figure. Images of genomes created with OGDRAW (Lohse et al. 2013).

Whole-genome alignment by Mauve (Figure 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 (Figure 2B). For example, the change of a number of tRNAs from an upstream position of *cox1* to a downstream position in several *Termitomyces* species probably occurred by successive inclusion and expulsion by the IR (Goulding et al. 1996).

A)



B)

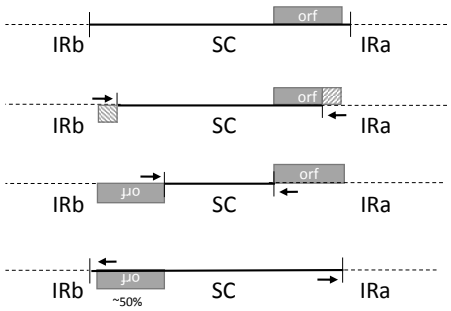


Figure 2: (A) Mauve alignment of four *Termitomyces* mitochondrial genomes: sp. T32 (used as reference), T159, T132 and T123. Coloured blocks indicate predicted homologies, which are connected through vertical lines. Genes associated with homologous regions are indicated in text for each block. If a 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 approximately 50% chance of the ORF translocating to a new position.

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). 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 (Figure 3).

G4DNA motif content is a divergent character across fungi, with most species we analysed showing relatively low frequencies of sequences with G4 potential (<0.5 per 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 *Termitomyces* species, as well as *B. zonata* and *T. rancida* all have estimates higher than 0.5 per 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 *Neurospora crassa*, to see if G4DNA was located differentially in these regions between species (Figure 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

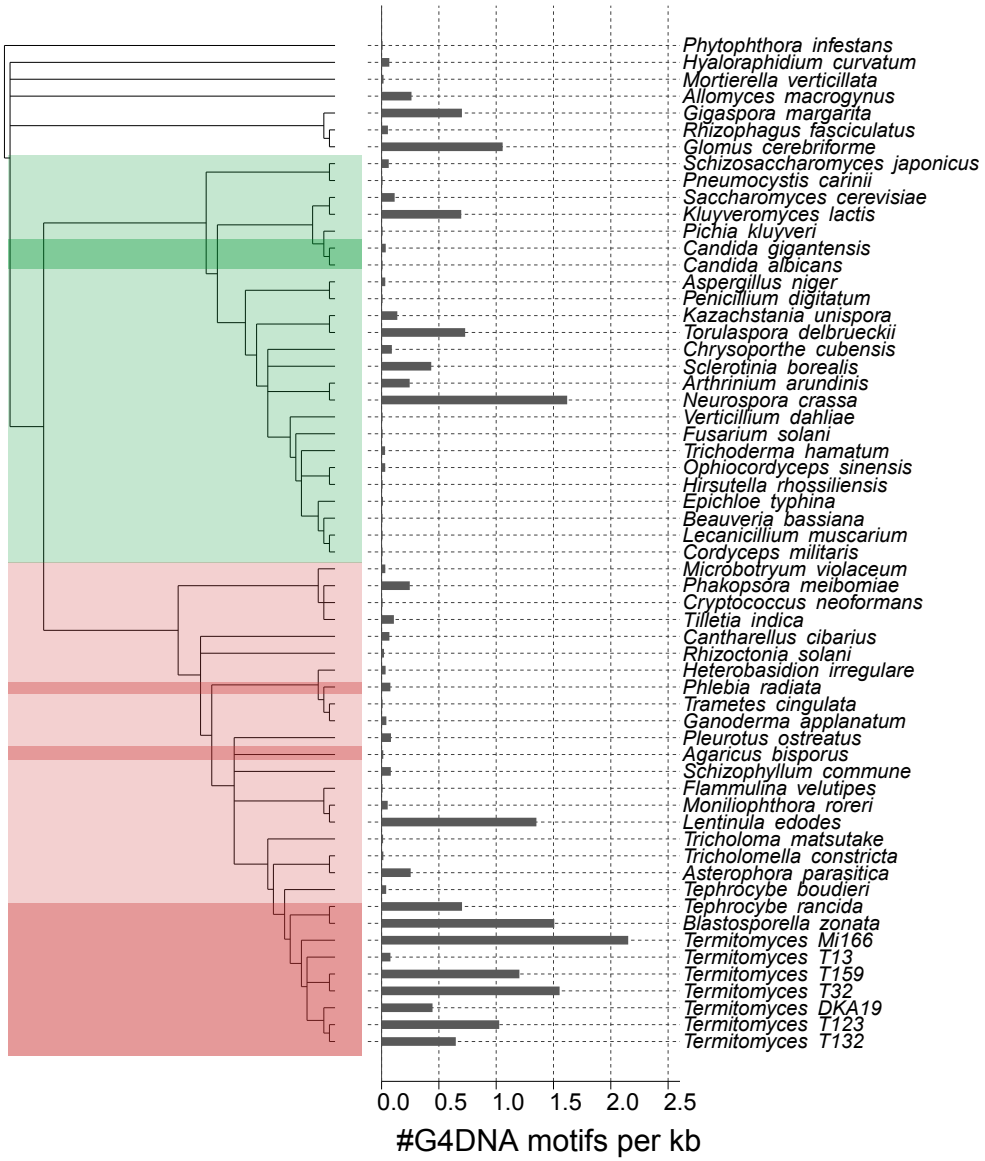


Figure 3: Barplot showing average G4DNA motif content per 1kb mtDNA for 62 species. Values were estimated with strict settings in G4Hunter (w=25, s=1.7). The phylogenetic tree was reconstructed using Maximum Likelihood with IQtree. Nodes with <70 bootstrap support were collapsed into polytomies. The 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.

ribosomal subunits, than expected by chance ($p<0.05$), whereas in human mtDNA there was no bias.

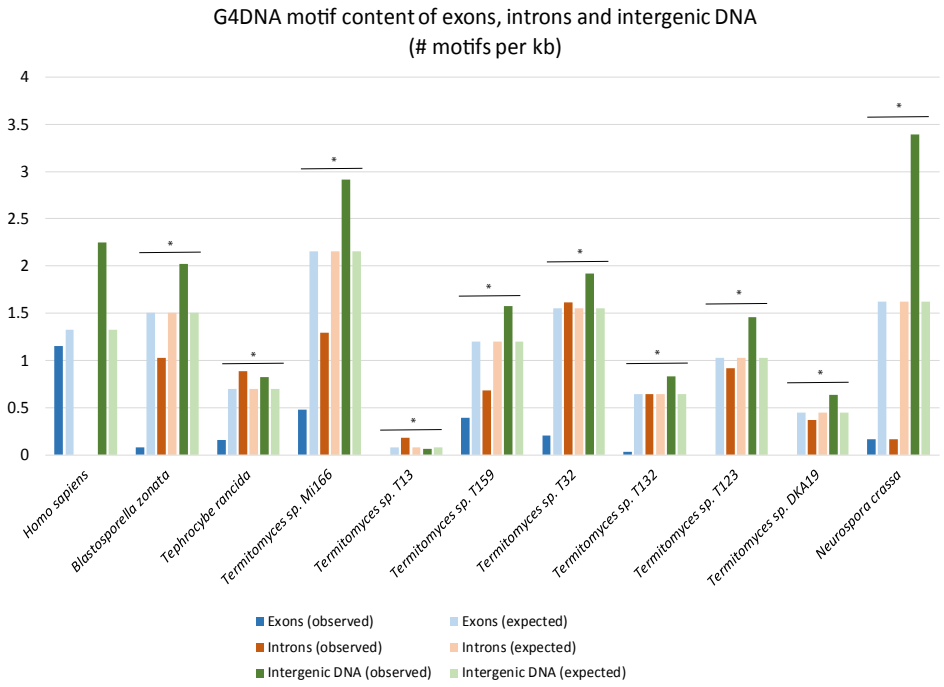


Figure 4: Observed and expected G4DNA motif content for exons, introns and intergenic regions in human and fungal mtDNA. Exons were only considered for the conserved protein-coding genes (cox1-3, nad1-6, atp6, atp8, atp9, cob and rps3) and the ribosomal subunits (rns and rnl, while these are not protein 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 significate deviation of observed values from the expected distribution (chi-square test, $p<0.05$).

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 this difference was very small (Figure 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 non-coding 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).

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. The highest value for each genome is shown in bold.

	Exons %	
Species	**SC**	**IR**
Blastosporella zonata	13.64	**15.50**
Tephrocybe rancida	18.91	**30.27**
Termitomyces sp. Mi166	13.20	**14.74**
Termitomyces sp. T13	15.52	**18.77**
Termitomyces sp. T159	14.37	**21.74**
Termitomyces sp. T32	**26.04**	24.85
Termitomyces sp. T123	18.91	**30.56**
Termitomyces sp. T132	16.58	**20.74**
Termitomyces sp. DKA19	**21.16**	5.18

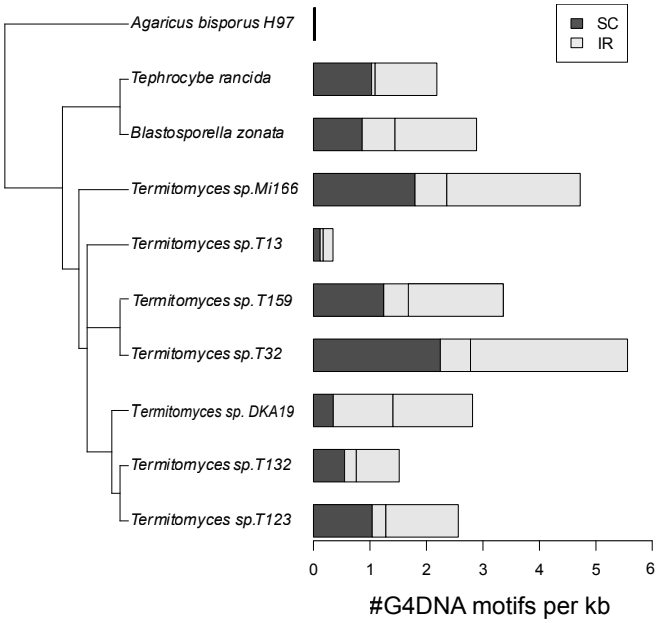


Figure 5: Average G4DNA motif content per 1kb mtDNA for Lyophyllaceae species with IRs, and *Agaricus* bisporus as outgroup. Dark grey denotes the G4DNA motif content for the SC region. Light grey indicates the G4DNA motif content of the IR. The expected 50/50 divide if the G4DNA motif

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.

Discussion

The mitochondrial genomes of *Termitomyces, T. rancida* and *B. zonata* are characterised 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 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 promotor 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 *Neurospora crassa, Termitomyces sp.* Mi166 and *Termitomyces sp.* T159, G4DNA occurs significantly less in introns than expected by chance (Figure 4).

Coding regions in fungal mtDNA almost never appear to contain G4DNA motifs (Figure 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 post-transcriptional 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 per kb) or even zero (Figure 3), but there are some notable exceptions to this trend. Fungi showing high frequencies of G4DNA motifs include the fungus *Neurospora crassa,* a model species for genetic and evolutionary studies. The only fungus for which mitochondrial G4DNA motif content was previously reported is *Saccharomyces cerevisiae* (Capra et al. 2010), and from our analysis this species appears to have a slightly enriched mitochondrial G4DNA motif

content compared to 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 in both cases mitochondria can function. That G4DNA can have an adverse effect on fitness is clear from its consistent absence from coding regions (Figure 4). As such it seems most likely that in those species where G4DNA is wholly absent selection against G4DNA also affects non-coding regions, or there is limited non-coding 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 & 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 *A. macrogynes*, only four overlapped G4DNA motifs. Similarly, more than a hundred GC-rich clusters are reported for the mtDNA of the yeast *Saccharomyces 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 (1kb+) mitochondrial inverted repeats, while rare in fungi, occur among others in *Candida*, *Agaricus*, *Phlebia radiata*, and *Agrocybe aegerita*. The IRs found in *Termitomyces, Tephrocybe rancida* and *Blastosporella 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 homologous recombination (Lambert et al. 2010). Consistent with this, G4DNA motifs are slightly but significantly more frequent within the IR than outside of it (Figure 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 homologous recombination (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 co-localization of replication origins in the repeats. The association of G4DNA with replication origins is supported in other organisms (Valton & Prioleau 2016; Valton et al. 2014; Paeschke et al. 2012), and the potential role of mitochondrial IRs in recombination-driven replication of mtDNA is supported by experimental results obtained from *Candida albicans* (Gerhold et al. 2010). In chloroplasts of maize, the IRs function as termini of linear monomers of ptDNA, and harbour the replication origins (Oldenburg & 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). The IR itself appears to show no reduced mutation rate in contrast to what was observed in some other studies (Maier et al. 1995; Palmer & Thompson 1982) (supplementary data 3). However, homologous recombination 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 behaviours: 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 (Figure 1) 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 centred 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 genomic 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 harbour 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 to the rest of the genome.

It is unclear how the IR originated. A small mitochondrial IR in *Agrocybe 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 terminal inverted repeats, 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 inverted repeat and an apparent increase in G4DNA content in the mitochondrial genomes of *Termitomyces*, *B. zonata* and *T. rancida*. We propose two non-mutually 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.

Acknowledgements

We thank Sabine Vreeburg for generating the RNA sequence alignment. We thank Ben Auxier for helpful comments on the draft version of this manuscript. We thank three

anonymous reviewers for their feedback. This study was funded by NWO-Vici grant 86514007.

References

Aanen DK et al. 2002. The evolution of fungus-growing termites and their mutualistic fungal symbionts. Proc. Natl. Acad. Sci. 99:14887–14892. doi: 10.1073/pnas.222313099.

Aldrich J, Cherney B, Merlin E, Williams C, Mets L. 1985. Recombination within the inverted repeat sequences of the Chlamydomonas reinhardii chloroplast genome produces two orientation isomers. Curr. Genet. 9:233–238. doi: 10.1007/BF00420317.

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. 117:33–43. doi: 10.1111/bij.12642.

Bankevich A et al. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. 19:455–477. doi: 10.1089/cmb.2012.0021.

Bedrat A, Lacroix L, Mergny JL. 2016. Re-evaluation of G-quadruplex propensity with G4Hunter. Nucleic Acids Res. 44:1746–1759. doi: 10.1093/nar/gkw006.

Bellanger JM et al. 2015. Plunging hands into the mushroom jar: a phylogenetic framework for Lyophyllaceae (Agaricales, Basidiomycota). Genetica. 143:169–194. doi: 10.1007/s10709-015-9823-8.

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. 289:29975–29993. doi: 10.1074/jbc.M114.567073.

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. doi: 10.1371/journal.pone.0102711.

Birky CW. 2001. The Inheritance of Genes in Mitochondria and Chloroplasts : Laws , Mechanisms , and Models. Annu. Rev. Genet. 125–48.

Blazier JC et al. 2016. Variable presence of the inverted repeat and plastome stability in Erodium. Ann. Bot. 117:1209–1220. doi: 10.1093/aob/mcw065.

Boetzer M, Henkel C V., Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 27:578–579. doi: 10.1093/bioinformatics/btq683.

Brien MAO, Misner I, Lane CE. 2014. Mitochondrial Genome Sequences and Comparative Genomics of Achlya hypogyna and Thraustotheca clavata. J. Eukaryot. Microbiol. 146–154. doi: 10.1111/jeu.12092.

Burger G, Gray MW, Lang BF. 2003. Mitochondrial genomes: Anything goes. Trends Genet. 19:709–716. doi: 10.1016/j.tig.2003.10.012.

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. doi: 10.1371/journal.pcbi.1000861.

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. http://

www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1934129.

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. 29:435–443. doi: 10.1093/bioinformatics/bts723.

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. doi: 10.1128/AEM.01815-17.

Court DA, Bertrand H. 1992. Genetic organization and structural features of Maranhar, a senescence-inducing mitochondrial linear plasmid of Neurospora crassa. Curr Genet. 22:385–397.

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. 2291–2299. doi: 10.1021/acs.biochem.6b00016.

Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Res. 1394–1403. doi: 10.1101/gr.2289704.

Dieckmann CL, Gandy B. 1987. Preferential recombination between GC clusters in yeast mitochondrial DNA. EMBO J. 6:4197–4203. doi: 10.1002/j.1460-2075.1987.tb02767.x.

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. 354:1067–1070. doi: 10.1016/j.bbrc.2007.01.093.

Du Z, Zhao Y, Li N. 2008. Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. Genome Res. 233–241. doi: 10.1101/gr.6905408.2.

Eddy J, Maizels N. 2006. Gene function correlates with potential for G4 DNA formation in the human genome. Nucleic Acids Res. 34:3887–3896. doi: 10.1093/nar/gkl529.

Frøslev TG, Aanen DK, Læssøe T, Rosendahl S. 2003. Phylogenetic relationships of Termitomyces and related taxa. Mycol. Res. 107:1277–1286. doi: 10.1017/S0953756203008670.

Fujitani Y, Yamamoto K, Kobayashi I. 1995. Dependence of frequency of homologous recombination on the homology length. Genetics. 140:797–809.

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. doi: 10.1038/srep28211.

Gerhold JM, Aun A, Sedman T, Jõers 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. 39:851–861. doi: 10.1016/j.molcel.2010.09.002.

Goulding SE, Olmstead RG, Morden CW, Wolfe KH. 1996. Ebb and flow of the chloroplast inverted repeat. Mol. Gen. Genet. 252:195–206. doi: 10.1007/s004389670022.

Grigoriev A. 1998. Analyzing genomes with cumulative skew diagrams. Nucleic Acids Res. 26:2286–2290. doi: 10.1093/nar/26.10.2286.

Gualberto JM et al. 2014. The plant mitochondrial genome : Dynamics and maintenance. Biochimie. 100:107–120. doi: 10.1016/j.biochi.2013.09.016.

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. 18:279–284. doi: 10.1038/nrm.2017.3.

Hofstetter V et al. 2014. Taxonomic Revision and Examination of Ecological Transitions of the Lyophyllaceae (Basidiomycota, Agaricales) Based on a Multigene Phylogeny. Cryptogam. Mycol. 35:399–425. doi: 10.7872/crym.v35.iss4.2014.399.

Huang W-C et al. 2015. Direct evidence of mitochondrial G-quadruplex DNA by using fluorescent anti-cancer agents. Nucleic Acids Res. 43:gkv1061. doi: 10.1093/nar/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 reeombination/circular restriction niap/oomycete). Proc. Natl. Acad. Sci. USA. 80:142–146. doi: 10.1073/pnas.80.1.142.

James TY et al. 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature. 443:818–822. doi: 10.1038/nature05110.

Johnson JE, Smith JS, Kozak ML, Johnson FB. 2009. In vivo veritas: using yeast to probe the biological functions of G-quadruplexes. Biochimie. 90:1250–1263. doi: 10.1016/j.biochi.2008.02.013.In.

Kearse M et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–1649. doi: 10.1093/bioinformatics/bts199.

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. doi: 10.1038/ncomms4216.

Lam T-W et al. 2013. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 1:1–6. http://f1000.com/718180400.

Lambert S et al. 2010. Homologous recombination restarts blocked replication forks at the expense of genome rearrangements by template exchange. Mol. Cell. 39:346–359. doi: 10.1016/j.molcel.2010.07.015.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701. doi: 10.1093/molbev/mss020.

Lang BF, Laforest MJ, Burger G. 2007. Mitochondrial introns: a critical view. Trends Genet. 23:119–125. doi: 10.1016/j.tig.2007.01.006.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9:357–9. doi: 10.1038/nmeth.1923.

Liachko I et al. 2014. GC-Rich DNA Elements Enable Replication Origin Activity in the Methylotrophic Yeast Pichia pastoris. PLoS Genet. 10. doi: 10.1371/journal.pgen.1004169.

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. doi: 10.1093/nar/gkt289.

Lynch M, Koskella B, Schaack S. 2006. Mutation pressure and the evolution of organelle genomic architecture. Science (80-.). 311:1727–1730. doi: 10.1126/science.1118884.

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. 614–628.

Maizels N. 2006. Dynamic roles for G4 DNA in the biology of eukaryotic cells. Nat. Struct. Mol. Biol. 13:1055–1059. doi: 10.1038/nsmb1171.

Mendoza O, Bourdoncle A, Boulé JB, Brosh RM, Mergny JL. 2016. G-quadruplexes and helicases. Nucleic Acids Res. 44:1989–2006. doi: 10.1093/nar/gkw079.

Miller MA, Pfeiffer W, Scwartz T. 2010. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. Proc. Gatew. Comput. Environ. Work. 1–8.

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. 32:268–274. doi: 10.1093/molbev/msu300.

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. 20:2619–2627. doi: 10.1111/j.1365-294X.2011.05090.x.

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. doi: 10.1186/1471-2148-14-121.

Oldenburg DJ, Bendich AJ. 2016. The linear plastid chromosomes of maize: terminal sequences, structures, and implications for DNA replication. Curr. Genet. 62:431–442. doi: 10.1007/s00294-015-0548-0.

Paeschke K, Capra JA, Zakian VA. 2012. DNA replication through G-quadruplex motifs is promoted by the S. cerevisiae Pif1 DNA helicase. Cell. 145:678–691. doi: 10.1016/j.cell.2011.04.015.DNA.

Palmer JD. 1983. Chloroplast DNA exists in two orientations. Nature. 301:92–93. doi: 10.1038/301092a0.

Palmer JD. 1985. Comparative Organization of Chloroplast Genomes. Annu. Rev. Genet. 19:325–354. doi: 10.1146/annurev.ge.19.120185.001545.

Palmer JD, Thompson WF. 1982. Chloroplast DNA Rearrangements Are More Frequent When a Large Inverted Repeat Sequence Is Lost. Cell. 29:537–550.

Paquin B, Laforest MJ, Lang BF. 2000. Double-hairpin elements in the mitochondrial DNA of allomyces: Evidence for mobility. Mol. Biol. Evol. 17:1760–1768. doi: 10.1093/oxfordjournals.molbev.a026274.

Peris D et al. 2017. Mitochondrial introgression suggests extensive ancestral hybridization events among Saccharomyces species. Mol. Phylogenet. Evol. 108:49–60. doi: 10.1016/j.ympev.2017.02.008.

Pietras Z et al. 2018. Dedicated surveillance mechanism controls G-quadruplex forming non-coding RNAs in human mitochondria. Nat. Commun. 9. doi: 10.1038/s41467-018-05007-9.

Rawal P et al. 2006. Genome-wide prediction of G4 DNA as regulatory motifs : Role in Escherichia coli global regulation Genome-wide prediction of G4 DNA as regulatory motifs : Role in Escherichia coli global regulation. Genome Res. 644–655. doi: 10.1101/gr.4508806.

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. doi: 10.1093/sysbio/sys029.

Schaffitzel C et al. 2001. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. Proc. Natl. Acad. Sci. 98:8572–8577. doi: 10.1073/pnas.141229498.

Shrivastav M, De Haro LP, Nickoloff JA. 2008. Regulation of DNA double-strand break repair pathway choice. Cell Res. 18:134–147. doi: 10.1038/cr.2007.111.

Smith DR, Keeling PJ. 2015. Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. Proc. Natl. Acad. Sci. 112:10177–10184. doi: 10.1073/pnas.1422049112.

Soares P et al. 2013. Evaluating Purifying Selection in the Mitochondrial DNA of Various Mammalian Species. PLoS One. 8. doi: 10.1371/journal.pone.0058993.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics. 25:1105–1111. doi: 10.1093/bioinformatics/btp120.

Turmel M, Otis Ch, 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. 96:10248–10253. doi: 10.1073/pnas.96.18.10248.

Valton AL et al. 2014. G4 motifs affect origin positioning and efficiency in two vertebrate replicators. EMBO J. 33:732–746. doi: 10.1002/embj.201387506.

Valton AL, Prioleau MN. 2016. G-Quadruplexes in DNA Replication: A Problem or a Necessity? Trends Genet. 32:697–706. doi: 10.1016/j.tig.2016.09.004.

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. 37:4194–4204. doi: 10.1093/nar/gkn1076.

Wernersson R, Pedersen AG. 2003. RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res. 31:3537–3539. doi: 10.1093/nar/gkg609.

Wilson AJ, Xu J. 2012. Mitochondrial inheritance: Diverse patterns and mechanisms with an emphasis on fungi. Mycology. 3:158–166. doi: 10.1080/21501203.2012.684361.

Wolters JF, Chiu K, Fiumera HL. 2015. Population structure of mitochondrial genomes in Saccharomyces cerevisiae. BMC Genomics. 16:1–13. doi: 10.1186/s12864-015-1664-4.

X. Xia and Z. Xie. 2001. DAMBE: Software Package for Data Analysis in Molecular Biology and Evolution. J. Hered. 92:371–373.

Xia X. 2012. DNA Replication and Strand Asymmetry in Prokaryotic and Mitochondrial Genomes. Curr. Genomics. 13:16–27. doi: 10.2174/138920212799034776.

Xu J, Zhang Y, Pun N. 2013. Mitochondrial recombination in natural populations of the button mushroom Agaricus bisporus. Fungal Genet. Biol. 55:92–97. doi: 10.1016/j.fgb.2012.09.004.

Yadav V kumar, 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. doi: 10.1093/nar/gkm781.

Yin S, Heckman J, RajBhandary UL. 1981. Highly conserved GC-rich palindromic DNA sequences flank tRNA genes in Neurospora crassa mitochondria. Cell. 26:325–332. doi: 10.1016/0092-8674(81)90201-4.

Zaug AJ, Podell ER, Cech TR. 2005. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. Proc. Natl. Acad. Sci. 102:10864–10869. doi: 10.1073/pnas.0504744102.

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. 209:1747–1756. doi: 10.1111/nph.13743.

Supplemental Data

1 - DNA isolation protocol for *Termitomyces*

Growing the material

1. Harvest the mycelium by scraping off the material. Try to pick up the least amount of agar possible. Transfer the material to safe-lock, 1.5 ml Eppendorf tubes together with 4-5 glass pearls for the bead beater.

Disrupting the cells and proteins

2. Crush the material by freezing with liquid N2 and using the bead beater for 10 seconds.
3. Repeat step 2.

- If the material does not appear to be disrupted enough (e.g. if large clumps of tissue are still present) repeat the procedure.

4. Add 500 µl CTAB (2% Hexadecyl.trimethyl.ammoniumbromide, 1.4M NaCl, 20mM EDTA, 100mM Tris.HCl pH=8) and 2 µl Proteinase K (15-20 mg/ml)
5. Put the Eppendorf tubes in the shaker and incubate them for 1 hour at 65 °C.

- Alternatively incubate them overnight at 55 °C

Phase separation - From here on work in the fumehood!

6. Add 500 µl Chloroform: Isoamylalcohol (24:1) and mix gently by inversion.
7. Centrifuge for 15 min at maximum speed and room temperature (still in the fumehood!).

- If the chloroform and water phase are not separated well enough centrifuge longer.

8. Carefully pipet off the water phase (top phase!) and transfer to a new Eppendorf tube. Do not be greedy; better to transfer less DNA then take some of the chloroform phase. Discard the lower (chloroform) phase to the proper waste container.

DNA precipitation

9. Add an equal volume (~300-400 µl) of ice-cold Isopropanol.

10. Put the tubes at -20 °C for a minimum of 20 minutes.
11. Centrifuge at maximum speed for 15 minutes at 4 °C
12. Remove the fluid from the pellet.

Washing DNA – from here on you can stop working in the fumehood

13. Add 300 µl 100% Ethanol to the pellet
14. Centrifuge at maximum speed for 5 minutes at room temperature
15. Remove the fluid from the pellet
16. Add 300 µl 70% Ethanol to the pellet

17. Centrifuge at maximum speed for 5 minutes at room temperature
18. Remove the fluid from the pellet
19. Dry the pellet in a vacuum for a couple of minutes

Collect DNA

20. Dissolve the dry pellet in 50 µl MQ water and store at -20 °C

2 – PCR protocol and gel pictures

PCR reaction mixture

5x GoTaq buffer	5 µl
25mM MgCl2	2 µl
10mM dNTP	1 µl
Forward primer	1 µl
Reverse primer	1 µl
GoTaq	0.1 µl
Template DNA 1:10	1 µl
MQ water	13.9 µl

Touchdown PCR program

Initial denaturation at **94°C** for **5min**

20 cycles (**60s** at **94°C**; **60s** at **60°C**; **60s** at **72°C**, decreasing annealing temperature by **0.5°C** after each cycle)

20 cycles (**60s** at **94°C**; **60s** at **55°C**; **60s** at **72°C**)

Final elongation at **72°C** for **10 min**

Primers

	Strain	Forward primer	Reverse primer
A	T132	TACCGCCATTGAACTCTAC	TTACGCGTGCAGATGCGCGT
B	T132	TTACGCGTGCAGATGCGCGT	AGTTGGTGATCTCCTTGAAG
C	T132	GGGGTTATTAATCACTGGAT	ACTTCTTTGGTATGGTGACC
D	T132	TTAAGGCCCTCCTTTCTGAT	ATCCAGTGATTAATAACCCC
H	T132	CACAGCTGAGTTAAGTCCTG	GGGGTTATTAATCACTGGAT
E	T13	GGGGTTATTAATCACTGGAT	TATGATTGCCTCCTACGGAG
F	T13	TATGATTGCCTCCTACGGAG	CCTTCTGTTATGACATCTCT
G	T13	CCGTACTTCGTACTTCGTAC	GCTTTAGTGGAAGACTCCGA
I	T13	CGAGATAGAGATTTCTAGGC	CCGTACTTCGTACTTCGTAC

Gel pictures:

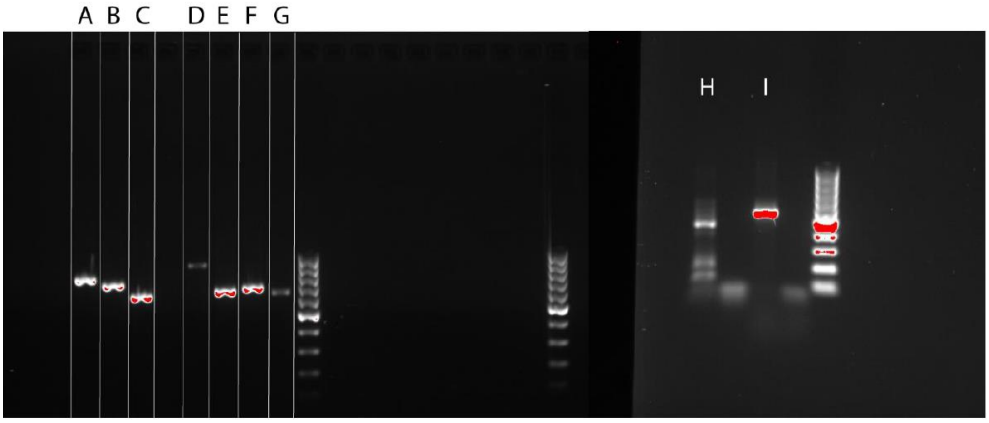


Figure 1: PCR products for all nine primer combinations. Lanes A,B,C,D and H correspond to the four IR/SC border regions in *Termitomyces* sp. T132. One border region was tested with two primer combinations to increase the chance of success as the sequence was predicted to cause problems for amplification. Lanes E,F,G and I correspond to the four border regions in *Termitomyces* sp. T13. All reactions yielded a product, confirming the presence of an inverted repeat in two locations in each mitochondrial genome.

3 - Mutation rates –Ka/Ks plots

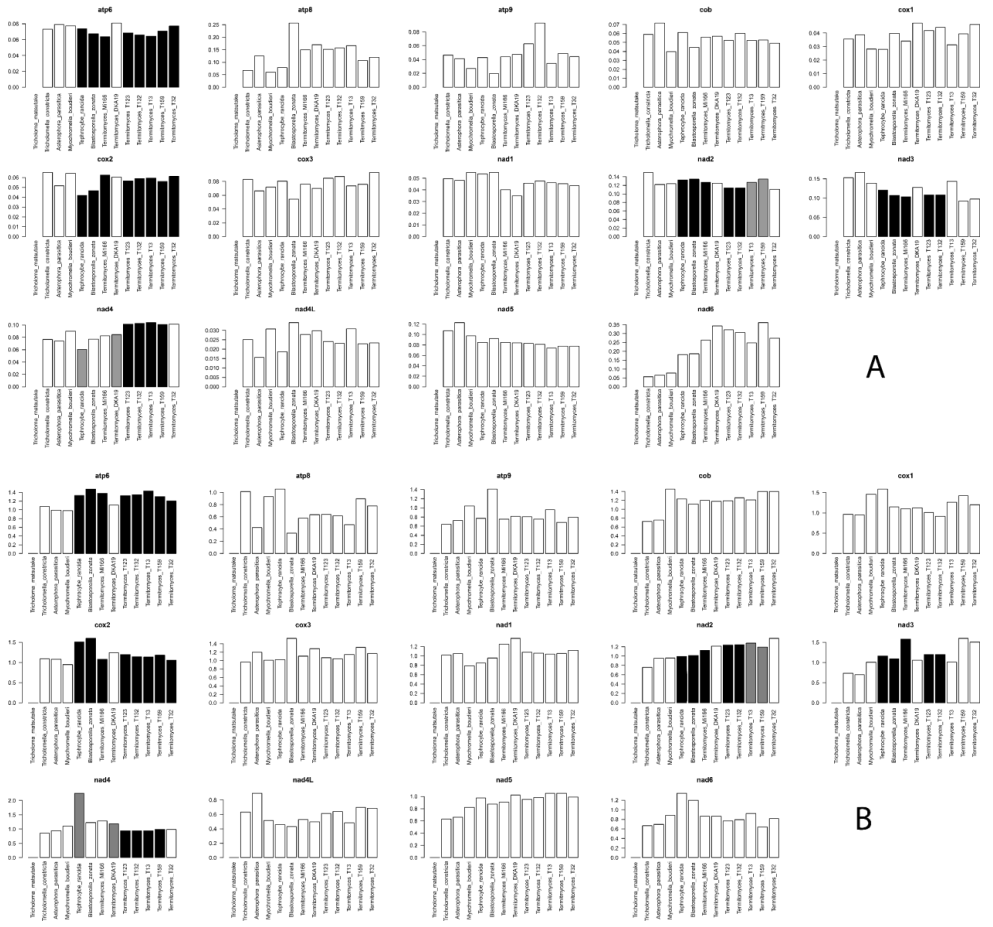


Figure 1: Bar charts showing Ka/Ks ratios (A) and Ks estimates (B) for each mitochondrial gene for each species sequenced for this study, with *T. matsutake* as outgroup. Genes contained within the mitochondrial inverted repeat are denoted by black bars, while genes partially within the IR are shown with grey bars. All genes appear to be subject to purifying selection (Ka/Ks < 1). Genes (partially) within the IR do not appear to have systematically higher or lower synonymous substitution rates than genes outside of it.

4.1 - G4DNA estimates

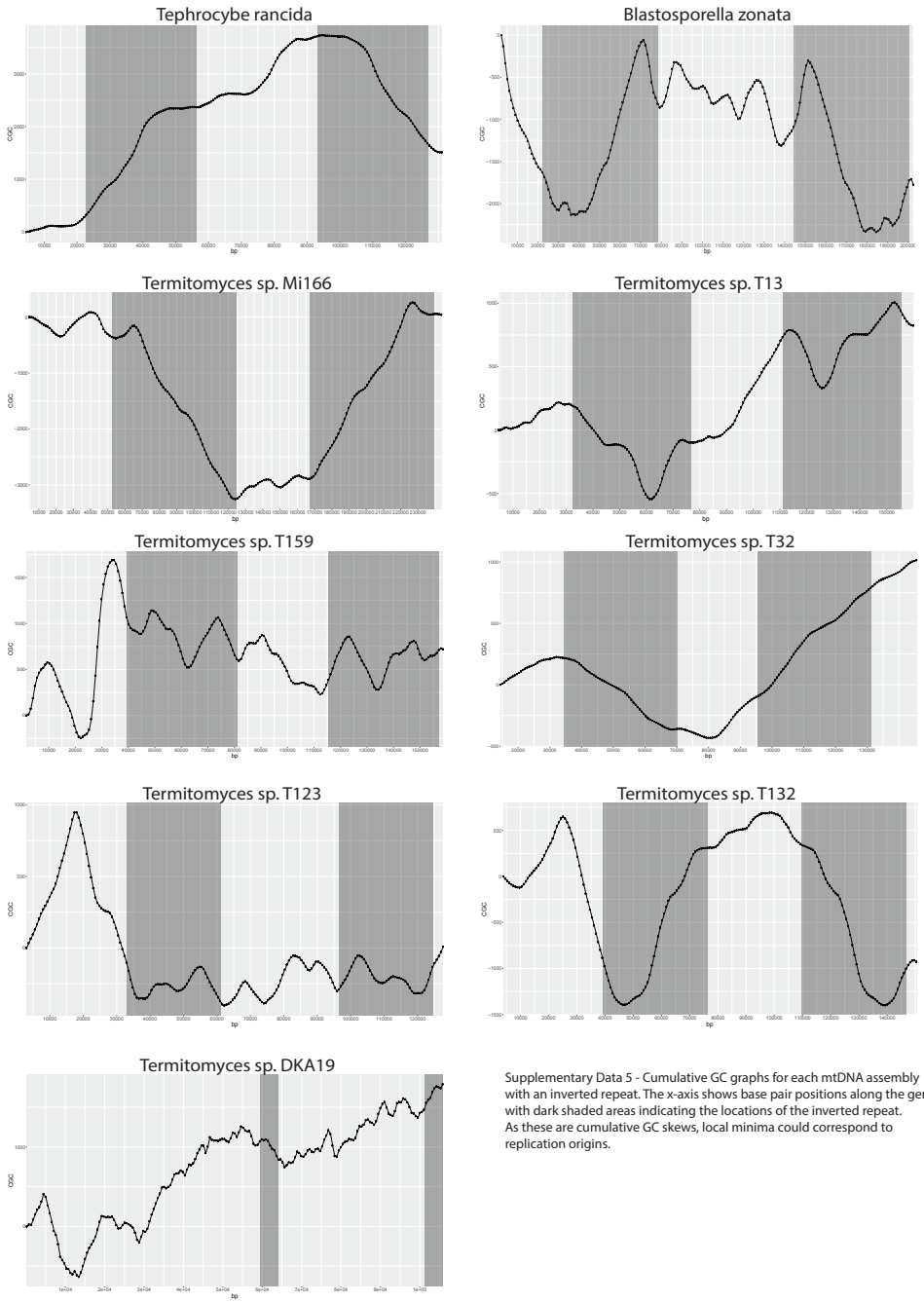
Taxon	Predicted G4DNA motifs (settings: w_25_s_1.7)	Predicted G4DNA motifs (settings: w_25_s_1.2)	mtDNA size (in bp)	Relative frequency predicted G4DNa motifs w_25_s_1.7	Relative frequency predicted G4DNA motifs w_25_s_1.2
Termitomyces_T32	204	823	131333	1.553303	6.266513
Termitomyces_T159	189	558	157156	1.202627	3.550612
Termitomyces_T13	12	122	155430	0.077205	0.784919
Termitomyces_Mi166	515	924	239317	2.151957	3.860988
Termitomyces_T132	95	241	146712	0.647527	1.642674
Termitomyces_T123	128	266	124711	1.026373	2.132931
Termitomyces_DKA19	47	136	105724	0.444554	1.286368
Blastosporella_zonata	302	476	200401	1.506979	2.375238
Tephrocybe_rancida	89	325	126794	0.701926	2.563213
Tephrocybe_boudieri	4	34	99774	0.040091	0.34077
Asterophora_parasitica	11	92	43328	0.253877	2.123338
Tricholomella_constricta	1	13	65087	0.015364	0.199733
Tricholoma_matsutake	1	6	76037	0.013151	0.078909
Pleurotus_ostreatus	6	29	73242	0.08192	0.395948
Agaricus_bisporus	2	21	135005	0.014814	0.15555
Lentinula_edodes	164	258	121394	1.350973	2.125311
Moniliophthora_roreri	5	38	93722	0.053349	0.405454
Flammulina_velutipes	0	6	88508	0	0.06779
Schizophyllum_commune	4	8	49704	0.080476	0.160953
Ganoderma_applanatum	5	18	119803	0.041735	0.150247
Trametes_cingulata	0	8	91500	0	0.087432
Phlebia_radiata	12	78	156348	0.076752	0.498887
Heterobasidion_irregulare	4	21	114193	0.035028	0.183899
Cantharellus_cibarius	4	21	58656	0.068194	0.35802
Rhizoctonia_solani	5	67	235849	0.0212	0.28408
Cryptococcus_neoformans	0	1	24919	0	0.04013
Phakopsora_meibomiae	8	38	32520	0.246002	1.168512
Tilletia_indica	7	22	65147	0.107449	0.337698
Microbotryum_violaceum	3	40	92107	0.032571	0.434278
Cordyceps_militaris	0	2	33277	0	0.060102
Lecanicillium_muscarium	0	2	24499	0	0.081636
Beauveria_bassiana	0	1	29961	0	0.033377
Epichloe_typhina	1	6	84630	0.011816	0.070897
Trichoderma_hamatum	1	2	32763	0.030522	0.061044
Hirsutella_rhossiliensis	0	3	62483	0	0.048013
Ophiocordyceps_sinensis	5	43	157539	0.031738	0.272948
Fusarium_solani	0	10	62978	0	0.158786
Verticillium_dahliae	0	1	27184	0	0.036786
Chrysoporthe_cubensis	8	33	89084	0.089803	0.370437
Neurospora_crassa	105	217	64840	1.619371	3.3467
Arthrinium_arundinis	12	56	48975	0.245023	1.143441

Sclerotinia_borealis	88	182	203051	0.433389	0.896327
Penicillium_digitatum	0	0	28970	0	0
Aspergillus_niger	1	6	31103	0.032151	0.192907
Kazachstania_unispora	4	19	29129	0.13732	0.652271
Saccharomyces_ cerevisiae	9	29	78917	0.114044	0.367475
Torulaspora_ delbrueckii	21	54	28793	0.729344	1.875456
Kluyveromyces_lactis	28	97	40291	0.694944	2.407486
Candida_albicans	0	15	33928	0	0.442113
Candida_gigantensis	2	60	55125	0.036281	1.088435
Pichia_kluyveri	0	10	43128	0	0.231868
Schizosaccharomyces_ japonicus	5	27	80059	0.062454	0.337251
Pneumocystis_carinii	0	7	26119	0	0.268004
Glomus_cerebriforme	63	259	59633	1.056462	4.343233
Rhizophagus_ fasciculatus	4	50	72251	0.055363	0.692032
Gigaspora_margarita	68	274	96998	0.701045	2.824801
Hyaloraphidium_ curvatum	2	27	29593	0.067584	0.912378
Mortierella_verticillata	1	16	58745	0.017023	0.272364
Allomyces_macrogynus	15	83	57473	0.260992	1.444156
Phytophthora_infestans	0	3	37957	0	0.079037

4.2 - G4DNA estimates of the inverted repeat

Taxon	Predicted G4DNA motifs in IR	Predicted G4DNA motifs in SC1	Predicted G4DNA motifs in SC2	Predicted G4DNA motifs in SC (both copies)	Size IR (in bp)	Size SC1	Size SC2	Size SC (combined)	relative frequency of G4DNA motifs in IR	relative frequency of G4DNA motifs in SC1	relative frequency of G4DNA motifs in SC2	relative frequency of G4DNA motifs in SC
Agaricus_ bisporus_H97	0	2	0	2	4559	91663	34224	125887	0	2.18E-05	0	1.59E-05
Blastosporella_ zonata	120	10	54	64	56326	22534	65545	88079	0.00213	0.000444	0.000824	0.000727
Termitomyces_ DKA19	11	18	7	25	4463	59447	37351	96798	0.002465	0.000303	0.000187	0.000258
Termitomyces_ Mi166	181	68	90	158	71831	52585	43096	95681	0.00252	0.001293	0.002088	0.001651
Termitomyces_ T123	37	32	22	54	28228	33020	35235	68255	0.001311	0.000969	0.000624	0.000791
Termitomyces_ T132	33	20	11	31	37211	39540	33112	72652	0.000887	0.000506	0.000332	0.000427
Termitomyces_T13	3	4	1	5	44175	32301	34779	67080	6.79E-05	0.000124	2.88E-05	7.45E-05
Termitomyces_ T159	63	20	42	62	41609	39591	34347	73938	0.001514	0.000505	0.001223	0.000839
Termitomyces_T32	66	29	43	72	35619	34443	25652	60095	0.001853	0.000842	0.001676	0.001198
Tephrocybe_ rancida	25	7	33	40	33754	22467	36819	59286	0.000741	0.000312	0.000896	0.000675

5 – Cumulative GC skews of mitochondrial genomes



Supplementary Data 5 - Cumulative GC graphs for each mtDNA assembly with an inverted repeat. The x-axis shows base pair positions along the genome, with dark shaded areas indicating the locations of the inverted repeat. As these are cumulative GC skews, local minima could correspond to replication origins.

Cumulative GC graphs for each mtDNA assembly with an inverted repeat. The x-axis shows base pair positions along the genome, with dark shaded areas indicating the locations of the inverted repeat. As these are cumulative GC skews, local minima could correspond to replication origins.

CHAPTER 3

3

Ancestral predisposition towards a domesticated lifestyle in the termite-cultivated fungus *Termitomyces*

Lennart JJ van de Peppel[1*], Mathijs Nieuwenhuis[1], Benjamin Auxier[1], Alexey A Grum-Grzhimaylo[2], Martha E C rdenas[3], Z Wilhelm de Beer[4], D Jean Lodge[5], Matthew E Smith[6], Thomas W Kuyper[7], Ana E Franco-Molano[8], Timothy J Baroni[9], Duur K Aanen[1*]

[1] *Laboratory of Genetics, Wageningen University & Research, Droevendaalsesteeg 1, 6708PB Wageningen, the Netherlands*

[2] *Microbial Ecology Department, Netherlands Institute of Ecology (NIOO-KNAW), Droevendaalsesteeg 10, 6708 PB Wageningen, the Netherlands*

[3] *Laboratorio de Micolog a y Fitopatolog a-LAMFU, Universidad de Los Andes, Carrera 1 # 10A-12, Bogot , Colombia.*

[4] *Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa*

[5] *Department of Plant Pathology, 2105 Miller Plant Sciences Building, University of Georgia, Athens, Georgia 30606*

[6] *Department of Plant Pathology, University of Florida, Gainesville, Florida 32611, USA*

[7] *Soil Biology Group, Wageningen University & Research, Box 47, 6700AA Wageningen, The Netherlands.*

[8] *Laboratorio de Taxonom a y Ecolog a de Hongos (TEHO), Instituto de Biolog a, Facultad de Ciencias Exactas y Naturales, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medell n, Colombia.*

[9] *Department of Biological Sciences, State University of New York, College at Cortland, PO Box 2000, Cortland, New York 13045, USA*

* *Corresponding authors (lennartvdpeppel@gmail.com, duur.aanen@wur.nl)*

Accepted for publication in Current Biology

Summary

The ancestor of termites relied on gut symbionts for degradation of plant material, an association that persists in all termite families[1,2]. However, the single lineage Macrotermitinae has additionally acquired a fungal symbiont that complements digestion of food outside the termite gut[3]. Phylogenetic analysis has shown that fungi grown by these termites form a monophyletic group the genus *Termitomyces*, but the events leading towards domestication remain unclear[4]. To address this, we reconstructed the lifestyle of the common ancestor of *Termitomyces* using a combination of ecological data with a phylogenomic analysis of 21 related non-domesticated species and 25 species of *Termitomyces*. We show that the closely related genera *Blastosporella* and *Arthromyces* also contain insect-associated species. Furthermore, the genus *Arthromyces* produces asexual spores on the mycelium, which may facilitate their dispersal by insects when growing on aggregated subterranean faecal pellets of a plant-feeding insect. The sister relationship between *Arthromyces* and *Termitomyces* implies that insect-association and asexual sporulation, which are present in both genera, preceded the domestication of *Termitomyces* and did not follow domestication as has been proposed previously. Specialisation of the last common ancestor of these two genera on an insect-faecal substrate is further supported by similar carbohydrate-degrading profiles between *Arthromyces* and *Termitomyces*. We describe a set of traits which may have predisposed the ancestor of *Termitomyces* towards domestication, with each trait found scattered in related genera and species outside of the termite-domesticated clade. This pattern indicates that the origin of the termite-fungus symbiosis may not have required large-scale changes of the fungal partner.

Key words: evolution, basidiomycetes, social insects, mutualistic symbiosis, domestication, fungus-growing termites, *Termitomyces*, mutualism, Lyophyllaceae, phylogeny

Results & Discussion

Phylogenetic relationships of the termitomycetoid clade

The family Lyophyllaceae (Basidiomycota), to which *Termitomyces* belongs, harbours species with diverse ecologies, including saprotrophic, parasitic and mutualistic lifestyles, with frequent transitions between them[5]. The genus *Termitomyces*, which engages in a mutualistic symbiosis with termites, presents an enigma on how such an intricate symbiosis could have evolved. Previous studies were unable to confidently identify the origin of this symbiosis due to a limited number of phylogenetic markers in the analyses or because closely related taxa were not included in the datasets[5-8]. To reveal the origin of the termite-*Termitomyces* symbiosis, using field collections and herbarium material we collected 39 samples of 11 genera spanning Lyophyllaceae. We obtained whole-genome sequences from these samples and added another seven publicly available assemblies. We reconstructed the phylogeny of these 46 taxa using 1131 conserved nuclear genes (Figure 1A). The topologies generated by both coalescent-based ASTRAL analysis (Figure S1) and concatenation-based IQ-TREE analysis were in agreement. Therefore, we will focus on the IQ-TREE phylogeny for our discussion. As high marker numbers inflate bootstrap support (BS)[9], we additionally calculated gene and site concordance factors (gcf and scf, Figure 1A)[10,11], which respectively show the fraction of gene trees or informative alignment sites supporting each node. Although concordance factors cannot be used to statistically test whether a node is well-supported by the data, they reflect the amount of ambiguity among loci for a given bifurcation.

Consistent with previous findings[5,8], we recover a well-supported termitomycetoid clade, which consists of the genera *Arthromyces*, *Blastosporella*, *Tephrocybe* and *Termitomyces* (Figure 1). Sister to the termitomycetoid clade is the genus *Myochromella*, although there is conflict among gene trees as indicated by the low gene concordance factor (36%). This indicates that the majority (64%) of gene trees do not support the quartet composition of this node, but it is the quartet supported by the largest share of gene trees. The sister group of the termitomycetoid-myochromelloid group is the clade consisting of the genera *Asterophora*, *Tricholomella*, *Sphagnurus* and *Lyophyllum*. The position of the genus *Calocybe* remains unresolved as we did not recover significant bootstrap support for this split. Our phylogenomic analyses provide strong support for the hypothesis that the sister group of *Termitomyces* is the genus *Arthromyces* (100BS, 72.1% gcf, 37.7% scf) (Figure 1). The sister group of the *Arthromyces-Termitomyces* clade is the clade consisting of two Asian species; *Tephrocybe* sp. 4 and *Tephrocybe* sp. 5, supported by 100% bootstrap, but showing some discordance among gene trees (43.1% gcf, 34% scf). The sister group of the *Arthromyces-Termitomyces-Tephrocybe* sp. 4-*Tephrocybe* sp. 5 clade is the clade represented in our dataset by the European and North American *T. rancida*, *Blastosporella zonata* from central and south America, *Tephrocybe* sp. 2 from South America and sp. 3

from Asia (100BS, 37.6% gcf, 33.1% scf). The increased discordance for these branches is not surprising as their branch lengths are relatively short, indicating a rapid diversification event. The deepest split in the termitomycetoid clade is represented by *Tephrocybe* sp. 1 from Australia and this relationship is well supported (100BS, 90.7% gcf, 45.7% scf).

Predispositions towards domestication in the ancestor of *Termitomyces*

The evolutionary steps in the three main groups of fungus-insect symbiosis remain unclear. For fungus-growing beetles, the accepted hypothesis is that domestication was contingent on dispersal by the beetles partner, as the wild relatives of the cultivated fungi are also dispersed by insects[12,13]. For fungus-growing ants it is proposed that either insect-facilitated dispersal or mycophagy was the initial step[14].The fungal partners in these two groups have multiple origins, and in the case of beetles also multiple origins of farmers, suggesting that fungal domestication in these groups was did not require many changes. In contrast, a single fungal lineage has been domesticated by a single group of termites[4], which makes the mutualism between termites and fungi a singularity . Such singularities can be interpreted as either difficult low-probability evolutionary events, or alternatively, that they are due to evolutionary priority effects, where first-movers suppress subsequent independent origins[15,16]. The biological data we collected and our phylogenetic reconstruction allowed us to identify a set of five traits shared by *Termitomyces* and the non-termite-associated sister group *Arthromyces*: a carbohydrate-degrading profile with a reduced potential to degrade plant-cell wall components, a rooting stipe (pseudorhiza), the formation of asexual spores (conidia), an insect-faecal association and the loss of clamp connections (Figure 1). Strikingly, these traits are shared to varying degrees by other members of the termitomycetoid taxa, suggesting that termitomycetoid fungi have a predisposition to domestication. Furthermore, some of these traits are found outside the termitomycetoid clade, such as the conidia-producing *Asterophora* or a (short) pseudorhiza in some specimens of *Tricholomella*[17,18]. We hypothesize that the combination of these traits in the ancestor of *Termitomyces* allowed it to colonize the comb formed by the most recent common ancestor of fungus-growing termites.

Reduced capacity for carbohydrate breakdown predates domestication

Fungi use a broad set of secreted carbohydrate-active enzymes (CAZymes) to breakdown and metabolise carbohydrates outside of their hyphal bodies. The CAZyme profiles of a species correlate with their ecology[19]. Previous research showed that *Termitomyces* has a reduced complement for the breakdown of these substances[20], but the timing of this reduction, whether predating the termite symbiosis or not, remained an open question. To test whether this reduction occurred pre- or post-domestication, we assessed the predicted CAZymes of the taxa in our dataset (Data S1). On average we identified 219 CAZymes per taxon (maximum: 375, minimum: 144). Species of *Termitomyces* have 197 CAZymes on average (max 251, min 144). The related non-termite associated taxa in

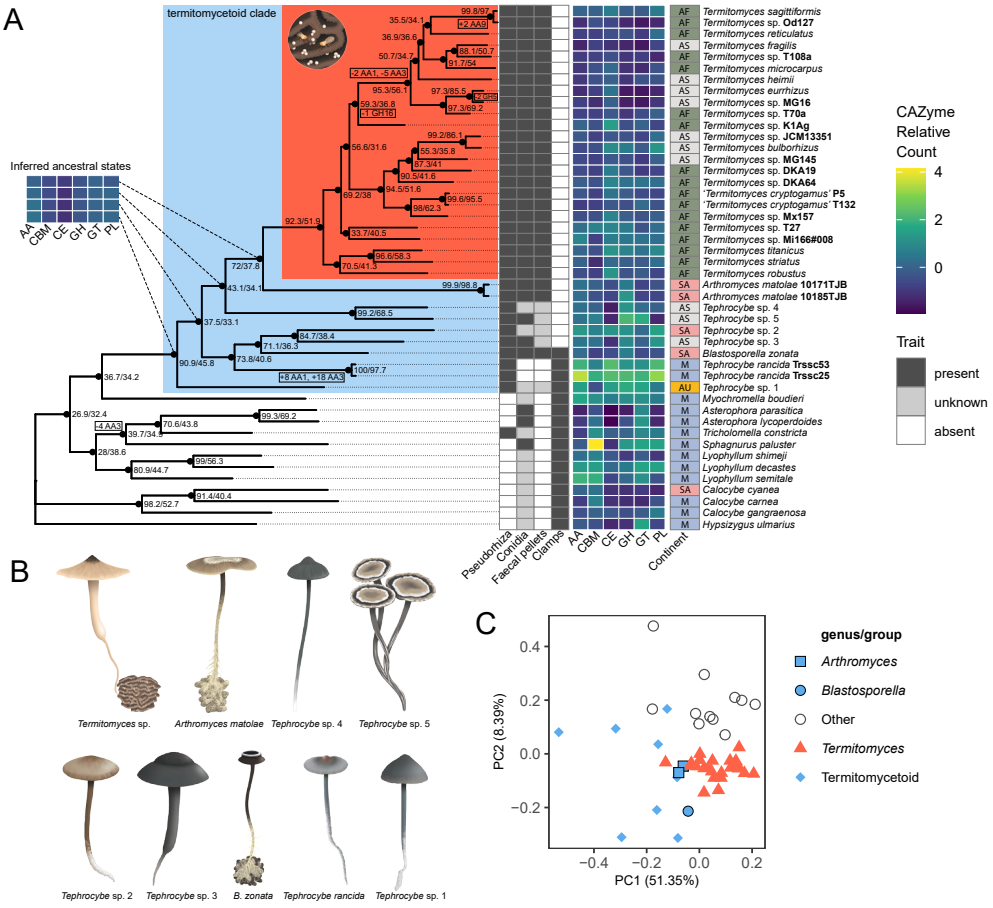


Figure 1: Phylogenomic and CAZyme analysis of the Lyophyllaceae showing that the insect-associated *Arthromyces* is sister to the genus *Termitomyces* and showing a reduction in CAZyme complement prior to the symbiosis. A: Maximum-likelihood tree based on 1131 core nuclear (BUSCO) genes. Numbers at the nodes represent gene concordance factor/site concordance factor, dots indicate 100% bootstrap support, numbers in boxes indicate significant changes in CAZymes. *Hypsizygus ulmarius* was used as an outgroup. Termitomycetoid clade highlighted in blue, termite-associated species highlighted in red. The most parsimonious reconstruction of the origin of the pseudorhiza is in the most recent common ancestor of the sister group of *Tephrocybe* sp. 1 as indicated. Matrix of morphological and CAZyme data: the left section of the matrix contains the character states of the predisposition traits: a pseudorhiza, production of asexual spores (conidia), an insect-faecal association, and the presence of hyphal clamp connections. The right section of the matrix shows a scaled heatmap of the total number of CAZymes split into the six main biochemical groups (Auxiliary Activity (AA), Carbohydrate Binding Module (CBM), Carbohydrate Esterase (CE), Glycoside Hydrolase (GH), Glycosyltransferase (GT) and Polysaccharide Lyase (PL). Colours indicate the difference of a taxon's CAZyme content from the average of the group, in standard deviations.

CAZyme data of selected ancestral nodes inferred by CAFE is indicated to the left of the phylogeny, using the same colour scale as the main matrix. Significant changes in CAZymes not explained by the phylogeny are indicated in boxes next to the nodes. The column on the right indicates the continent on which each species occurs; Australia (AU), Africa (AF), Asia (AS), South America (SA) or multiple continents (M). B: drawings of representative species in the termitomycetoid clade showing their overall morphology and the presence and/or absence of the predisposition traits. C: Principal component analysis of the same CAZyme data as in the CAZyme matrix except that the six main groups are split into 119 CAZyme subfamilies, numbers in parentheses indicate the percentage of variance explained by a principal component. Note that samples of *Termitomyces* (red triangles) and *Arthromyces* (blue squares) cluster together.

the genera *Arthromyces* and *Blastosporella* have a slightly higher CAZyme complement, with 233 and 232 CAZymes, respectively. There is no clear pattern of change related to *Termitomyces* in all six functional CAZy classes (Figure 1A). Principal component analysis of a finer separation of CAZymes revealed that generally *Termitomyces* species cluster together, with *Blastosporella* and *Arthromyces* nearby, despite the large genetic distance between *Termitomyces* and *Blastosporella* (Figure 1C). Correcting the principal component analysis for phylogeny[21] did not reveal any source of CAZyme variation from other members of the Lyophyllaceae (Figure S2E). Variation in CAZyme sets is unlikely to be related to sequencing quality, as we found no correlation between the number of reads used in the assemblies and the CAZyme, BUSCO, or the entire predicted proteome content (Figure S2A-D).

Analysis of the evolutionary history of the CAZyme gene families revealed no significant changes that could be ascribed to the transition to the termite symbiosis, which indicates that the reduced CAZyme complement predates the symbiosis (Figure 1). Five gene families of the 119 total had evolutionary histories that were not explained by the phylogeny alone as tested using the CAFE analysis: AA1, AA3, AA9, GH16 and GH5. The changes in the evolutionary history of AA1 and AA3, both used in the oxidation of lignin, were based on increased gene family number in the two *Tephrocybe rancida* strains. The changes in the GH5 and AA9 families, both families involved in cellulose degradation, while significant, were restricted to individual species, with no internal nodes showing unexpected changes. The GH16 family, involved in breaking various β-1,3-glucan bonds, showed changes at the tips of the tree, as well as a reduction within *Termitomyces*.

Insect-faecal associations predate domestication

During field work, collections of *A. claviformis*, *A. matolae* and *B. zonata* showed that these species were associated with aggregated clumps of insect-faecal pellets in all cases (Figure 2). Faecal pellets upon which *A. matolae* was growing were composed of woody plant material (Figure S3). We were unable to identify the insect source of the pellets.

To identify the source of the faecal pellets in our *B. zonata* samples, we collected nearby abundant beetle larvae whose fresh faecal pellets were macroscopically similar to those found with *B. zonata* samples (Figure 2D). We used DNA barcode sequences to identify these as larvae of a Scarab beetle (Scarabaeidae), although we could not identify them to the species level, due to a lack of an identical or close matches in NCBI Genbank (STAR methods). Our novel finding of the insect-faecal pellet substrate of *Arthromyces* and *B. zonata,* raises the question whether other species in the termitomycetoid clade are also associated with insect faeces. We were unable to re-collect any of the species of *Tephrocybe* except *T. rancida*, but despite extensive efforts, we found no indication that *T. rancida* grows on insect faeces (STAR Methods and Figure S3). This apparent lack of an insect-faecal relationship is supported by the higher total number of CAZymes in the *T. rancida* genome compared to *B. zonata* (Data S1), including significant increases in AA1 and AA3 (Figure 1A). The phylogenetic position of the five unidentified species of *Tephrocybe* in the termitomycetoid clade, the pseudorhiza and their similarity in CAZyme profiles warrants re-collection and detailed study of the substrate. It remains unclear if these taxa have unrecognized associations with insect faeces.

Conidial production predates domestication

While all mushroom-forming fungi produce sexual spores (basidiospores) for reproduction, only some species produce asexual spores (conidia), either on the mycelium or in rare cases on the mushroom[22]. However, several species in the termitomycetoid clade, produce conidia (Figure 1). In species of *Termitomyces* only the mycelium produces conidia[23] (Figure 3A-D), which are ingested by the termites and mixed in the gut to inoculate fresh fungus combs[24]. These conidia are produced both inside termite mounds as well as when grown in laboratory culture. Previously, it has been reported that *Arthromyces* species produce conidial chains by fragmentation of terminal hyphae (arthroconidia) on the entire mushroom while *B. zonata* produces ornamented conidia in small spore heads by a budding process (so-called blastoconidia) on the mushroom cap[25]. Here, we report that *A. matolae* also produces dikaryotic conidia on the mycelium in laboratory culture (Figure 3E-H) (Figure S4A). These conidia are encased in an elaborate structure composed of a hyphal net with large extended setae composed of single cells. We further found that *B. zonata* produces both dikaryotic blastoconidia and arthroconidia on the mycelium in culture (Figure S4B-D). Arthroconidia were also detected on the mushrooms of *Tephrocybe* sp. 3 and *Tephrocybe* sp. 5. We did not find conidia in the following taxa: *T. rancida*, *Tephrocybe* sp. 1, *Tephrocybe* sp. 2 and *Tephrocybe* sp. 4. The current lack of laboratory cultures for the various *Tephrocybe* spp. prevents the confident pinpointing of the origin, or origins and subsequent losses, of conidial production in the termitomycetoid clade. However, the most parsimonious reconstruction is that the common ancestor of *Arthromyces* and *Termitomyces* produced conidia.

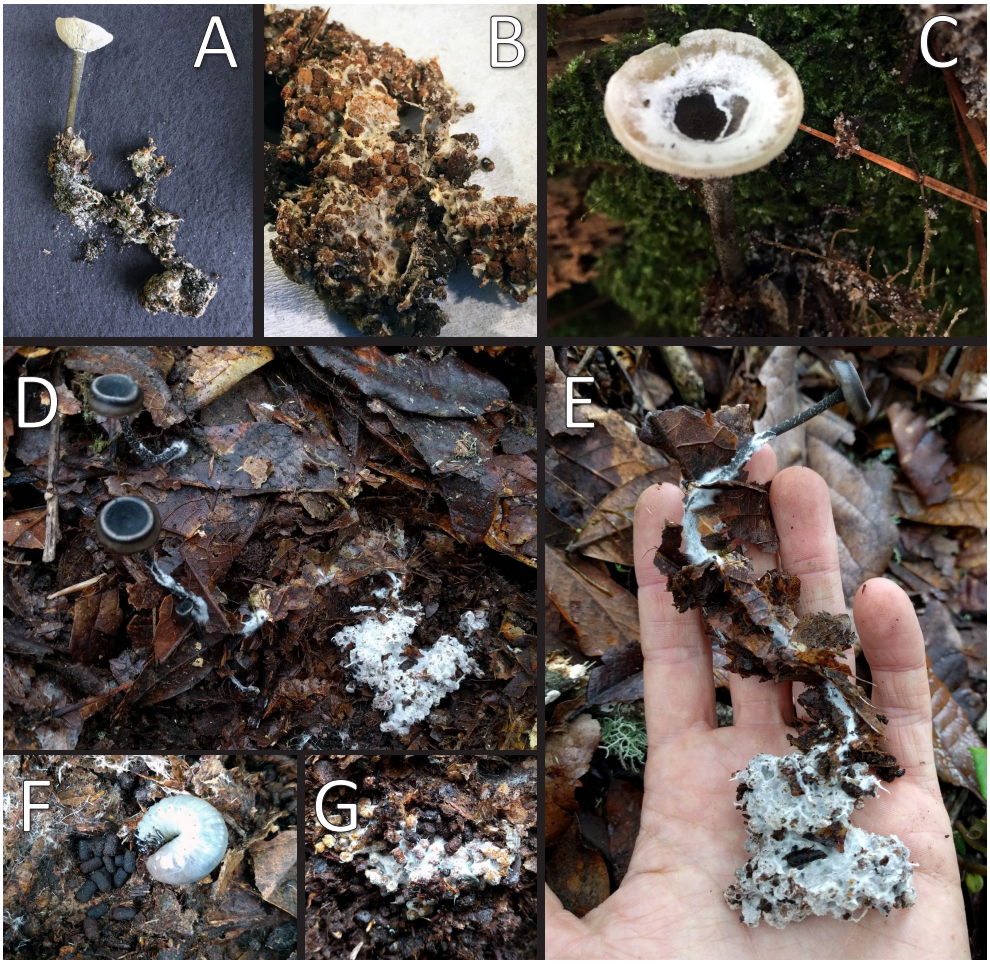


Figure 2: Insect-faecal associations found in *Arthromyces* matolae and *Blastosporella zonata*. A: Mushroom of *A. matolae* connected to a faecal pellet mass. B: Close-up of the faecal pellets. C: Pileus of *A. matolae* showing the dark arthroconidia. D: *B. zonata* mushrooms and pellet mass. E: Detail of the mushroom of *B. zonata* showing the attachment with the pseudorhiza to the pellet substrate. F: Beetle larva next to fresh faecal pellets. G: Close-up of the faecal pellet mass showing white rot in the pellets.

A rooting stipe predates domestication

All species within the termitomycetoid clade are able to produce a rooting stipe (pseudorhiza). Within *Termitomyces* , *T. microcarpus* may only produce a tiny pseudorhiza in some cases, probably as a response to epigeous fruiting on expelled comb material[26]. The pseudorhiza is a specialized part of the stipe which pushes the immature mushroom of the fungus from the buried subterranean nutrient substrate to the soil surface[27] and has evolved independently several times in Basidiomycota. The nutrient substrate can be a

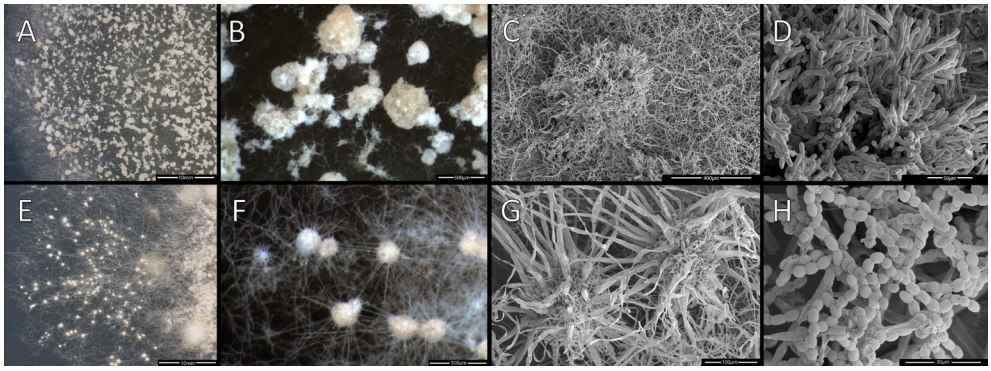


Figure 3: Macroscopic and SEM microscopic images of the conidiophores of *Termitomyces cryptogamus* and *Arthromyces matolae*. Conidiophores in laboratory culture of *T. cryptogamus* (P5) (A,B). Electron microscope image of a conidiophore showing the chains of conidia (C). Close-up of a conidiophore showing the branched chains of conidia of *T. cryptogamus* (D). Laboratory culture of *A. matolae* (FLAS-F-62734) showing the conidiophores (E,F). Electron microscope image of *A. matolae* showing two conidiophores with large aerial setae (G). A close-up of an *A. matolae* conidiophore showing the chains of arthroconidia (H).

tree root in the case of *Phaeocollybia* and *Xerula radicata*[28,29], a wood mouse or mole latrine in the case of *Hebeloma radicosum*[30,31], and insect faeces in the case of *Arthromyces*, *B. zonata* (Figure 2) and *Termitomyces*, where insect faeces form a specialised fungus comb. All termitomycetoid species are able to produce a pseudorhiza (Figure 1) which indicates a single transition towards growth on a buried nutrient substrate in this group.

Loss of clamp connections predates domestication

There is a striking pattern in a range of mutualisms with convergent increases in genome copy numbers per cell, either through polyploidy or multiple nuclei[32]. Examples are the multinucleate cells of the fungi cultivated by leaf-cutting ants[33], the fungi forming arbuscular mycorrhizae with plants[34], the polyploid endosymbiotic plastids and mitochondria of eukaryotic cells[35], and even our own domesticated crops, most of which are polyploids[36]. *Termitomyces* also fits in this pattern with multinucleate cells containing up to 10 nuclei, lacking the specialised morphological structures, clamp connections, that maintains nuclei per cell in most other basidiomycete fungi[32,37]. This striking similarity suggests that increased ploidy of symbionts may be selected as a consequence of a symbiotic lifestyle. However, the most parsimonious reconstruction is that clamp connections were lost before domestication of *Termitomyces* since its sister group *Arthromyces* also lacks clamp connections[25]. The origin of multinucleate cells without clamp connections presumably is even older, predating the split between the *Termitomyces-Arthromyces* clade and the *Tephrocybe* sp. 4 and sp. 5 clade, but following the second split in the termitomycetoid

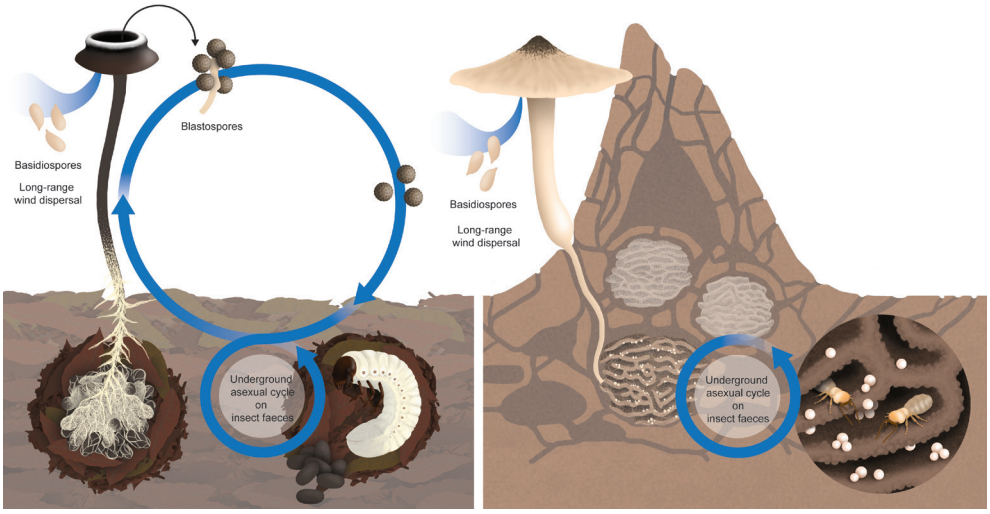


Figure 4: Proposed insect-fungal interaction model of insect-faecal associated mushroom-forming fungi compared to the lifecycle of the domesticated genus *Termitomyces*. In the aboveground part, asexual spores produced on the mushroom are used for local dispersal and rapid inoculation of the faecal pellets. In the belowground part, asexual spores may stick to an insect and inoculate fresh batches of faecal pellets. Additionally, the conidia could also facilitate faster colonization by reinoculation of the substrate via the insect host. Wind-dispersed basidiospores are used for long-range dispersal and sexual recombination.

clade, between *Tephrocybe* sp. 1, which has clamp connections, and its sister group, most members of which lack clamp connections.

Origin of termite-fungus symbiosis

Three pathways have been described to explain the evolution of a domestication interaction; commensal, prey-or-harvest, and direct[38,39]. The latter two primarily apply to human domesticators since they involve conscious selection of a potential domesticate. In the commensal pathway, a two-way partnership can arise if species A enters the habitat of species B where it can make use of a niche provided by species B[39]. Previously, the sister group of *Termitomyces* and its biology have remained unknown, preventing identification of this potential niche. The consumption of fungus-infested wood has been shown to increase survival in several different species of termites[40]. Some species of termites are also known to consume mushrooms[41] and fungus-growing termites have been observed to consume mushrooms other than *Termitomyces*[42]. The observation that termites consume and are attracted to rotting wood led to the hypothesis that mycophagy was the initial step towards domestication[14,43,44]. This explanation may be too simplistic as it does not recognize a potential niche provided by the termites that could be utilized by the fungus. Our novel findings on the biology of species in the termitomycetoid clade suggests that

the ancestor was not a generalist wood degrader but already possessed a suite of traits which predisposed it towards domestication. This is unique among the fungus-growing insects as ancestral predispositions have not been documented in fungus-growing ants and ambrosia beetles.

Our phylogenetic analysis provides strong support for a sister-group relationship between *Arthromyces* and *Termitomyces*. Both genera share the five traits discussed above, so the most parsimonious reconstruction is that the common ancestor of *Termitomyces* grew on insect faeces and had the carbohydrate-degrading enzymatic profile to utilize that growth substrate, produced conidia, had multinucleate cells with no clamp connections and had a rooting stipe. Our results therefore imply that an insect-faecal association predated the termite-fungus symbiosis. This insect-faecal association may even predate the common ancestor of *Arthromyces* and *Termitomyces* as several other taxa in the termitomycetoid clade exhibit various combinations of these five traits.

All non-Termitid termite families rely on cellulolytic protist gut symbionts for the digestion of wood[45]. However, these gut symbionts have been lost in the Termitidae[46,47]. The loss of gut symbionts was correlated with a diversification of feeding habits, including fungus-farming, bacteria-farming and soil-feeding[2,48]. Recent evidence suggests that the subfamily Sphaerotermitinae, the only extant member of which constructs combs similar to fungus-growing termites but are instead colonized by bacteria, is the sister group of the fungus-growing termites[49,50]. The most parsimonious reconstruction for comb evolution is that comb-building was present in the common ancestor of the Sphaerotermitinae and the fungus-growing termites[50]. After the loss of the flagellated protists, the ancestral comb may have served as an external rumen in fungus-growing termites[3], providing a suitable substrate for a fungus adapted to grow and reproduce on a similar faecal substrate. As wood-decay fungi are more efficient lignin degraders compared to lignocellulolytic bacteria[51], the increased capacity to degrade lignin and cellulose combined with a reduced capacity to break down oligosaccharides of the ancestors of *Termitomyces* may have increased the amount of useable calories in the comb substrate which directly benefitted the termites. Given the sister-group relationship between *Arthromyces* and *Termitomyces,* the most parsimonious reconstruction is that the appendiculate conidiophores formed by *Arthromyces* and the nodules of *Termitomyces* are homologous. The conidia may have served an important role in maintaining the fungus as the dominant species within the ancestral comb since the conidia could facilitate local dispersal and continuous substrate reinoculation (Figure 4).

The conidia produced on the mushroom in *Arthromyces* and several other species in the termitomycetoid clade may be involved in local dispersal since they are thick-walled and this likely increases their chances of survival during periods when substrate is unavailable.

This strategy has been suggested for the mycoparasitic genus *Asterophora* as well because their hosts are only available seasonally and patchily[17,52]. It seems likely that they serve the purpose of survival structures such as chlamydospores or sclerotia that are found in a wide array of Basidiomycota[22,53].

According to our hypothesis, the niche provided by the termites for the ancestor of *Termitomyces* was the pre-digested plant material in the ancestral comb. Besides mineral nutrients obtained from consuming the fungus, the termites could also benefit from the increase in available calories of the comb as a result of fungal degradation and the oligosaccharides left behind by the fungus. The fungus benefitted from a physically pre-treated substrate with an enlarged surface area combined with local dispersal and reinoculation by the termites.

Non-wind dispersed asexual spores (produced by the fungus) that are locally dispersed by a different species (the termites) is analogous with human domestication and propagation of food plants. It has been proposed that the reduction of natural seed dispersal, which includes the loss of seed shattering, is a key trait selected for under cultivation[54,55]. Most seed crops have non-shattering seeds, which renders a plant species with limited dispersal abilities and primarily dependent on humans for survival and propagation[56]. Similarly, *Termitomyces* depends on termites for substrate colonisation and accordingly for survival and propagation[57]. The finding that the sister group of *Termitomyces* produces conidia and is associated with insect faeces lends support to the hypothesis that one of the first steps in the domestication process was local dispersal by insects.

Acknowledgements

We would like to thank the Colombian students from the University of Antioquia who assisted with collecting *B. zonata* during fieldwork. We thank Dirk Stubbe for providing a specimen of *Tephrocybe* sp. 5 We thank Marc Maas for creating Figure 4. We thank the Royal Botanical Gardens KEW herbarium for letting us use their facilities and providing specimens. We thank the field mycologists who helped us with taxon sampling; Carolien Reindertsen, Mirjam Veerkamp, Arthur Grupe, Jenny Rogers, Jacob Kalichman and Gert Immerzeel. We thank Natascha Oosterwijk for making some of the microscopic images. We thank Marcel Giesbers from the Wageningen Electron Microscopy Centre for his help in making the electron microscopy images. DKA, LJJvdP, MN, AAGG and BA were supported by the Netherlands Organisation for Scientific Research (DKA, LJJvdP, MN, and AAGG by VICI: NWO 86514007; DKA and BA by ALWGR.2017.010). TJB and DJL acknowledge grants from NSF (DEB-9525902 and DEB-0103621 Belize, Puerto Rico, Dominican Republic), TJB for support from the National Geographic Society (Belize) and the New York Botanical

Gardens (Thailand and Australia) supporting field research that allowed discovery of many of the novel non-*Termitomyces* taxa included in this study.

The Colombian biological material used in this work was collected under the permission given to the University of Antioquia according to resolution 0524 of May 27, 2014 issued by the National Environmental Licensing Authority - ANLA.

Author contributions

Conceptualization, LJJvdP, DKA; Methodology, LJJvdP, DKA, MN, BA; Investigation, LJJvdP, DKA, MN, AAGG, TJB, MES, MEC AEFM, TWK, ZWdB; Formal analysis; LJJvdP, MN, BA; Resources, TJB, MES, MEC, AEFM, DJL, TWK; Data Curation, MN; Writing Original Draft, LJJvdP. All of the authors commented on the first draft and approved the final version of the manuscript.

Declaration of interests

The authors declare no competing interests.

STAR★Methods

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Lennart van de Peppel (lennartvdpeppel@gmail.com)

Materials availability

A culture of *Arthromyces matolae* (FLAS-F-62734) was submitted to the culture collection of the Westerdijk Fungal Biodiversity Institute under the accession number CBS 147616.

Data and code availability

Newly generated raw reads of the fungal taxa are available in the Sequence Read Archive (SRA) with accession numbers: SRX10313000-SRX10313007, SRX10337354-SRX10337371 and SRX4910404-SRX4910415. Assembled genomes are available at NCBI Genbank with accession numbers: GCA_017580835.1, GCA_017607575.1, GCA_017657195.1, GCA_017657225.1, GCA_017657235.1, GCA_017657275.1, GCA_017657295.1, GCA_017657315.1, GCA_017657335.1, GCA_017657355.1, GCA_017657375.1, GCA_018220975.1, GCA_018221615.1, GCA_018221635.1, GCA_018221655.1, GCA_018221735.1, GCA_018221785.1, GCA_018221805.1, GCA_018282005.1, GCA_018282025.1, GCA_018849495.1, GCA_018850235.1, GCA_018850255.1, GCA_018850275.1, GCA_018850815.1, GCA_018851285.1, GCA_018851305.1, GCA_018851325.1, GCA_018851835.1, GCA_018854895.1, GCA_018855395.1, GCA_018855915.1, GCA_018856295.1, GCA_018857265.1, GCA_018857285.1, GCA_018857305.1, GCA_018858115.1. Beetle larva sequences are available at NCBI Genbank with accession numbers: MW698941 and MW698942.

Scripts used for filtering contigs and removing bacterial contamination from the assemblies in the bioinformatic analyses as well as the scripts used in the CAZyme analysis are available at: https://github.com/BenAuxier/Termite.Domestication

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We sampled a total of 47 taxa, of which 25 were *Termitomyces*, across the entire family of the Lyophyllaceae and within the Lyophyllaceae we focused primarily on the termitomycetoid clade[5]. For the three remaining clades we sampled one to four representative species. We collected specimens from various sources; the *Termitomyces* samples were obtained from our in-house culture collection, the Royal Botanical Gardens KEW herbarium and from mushrooms stored in ethanol collected in Ivory Coast. Other non-*Termitomyces* lyophylloid herbarium specimens were obtained from the Royal Botanical Gardens

KEW herbarium and the Cort herbarium. Fresh Lyophylloid mushrooms were provided by various collectors from the Netherlands, DNA was isolated either directly from these mushrooms or from tissue cultures. Tissue cultures were made by cutting the stipe or pileus and moving a sterile piece of tissue from the inside of the stipe or pileus with sterile forceps into a petri dish containing malt yeast extract agar (per litre demi water: 20 g malt extract, 2 g yeast extract, 15 g agar) and streptomycin (30mg/L) against bacterial contamination. A full overview of all samples can be found in Data S2.

METHOD DETAILS

Detection and collection of faecal pellets

Mushrooms of *Blastosporella zonata* and pellets were collected in Murillo - Tolima, Colombia. Between ten and 15 pellets were stored in 1.5ml Eppendorf tubes with pure ethanol for subsequent analysis. In some cases, pellets were found very close to the pellet mass on which *B. zonata* was growing. These pellets, which did not show fungal colonization, were collected separately.

On two occasions (collection Bzo6 and Bzo8) scarabid beetle larvae were found in close proximity of the pellet substrate. A total of six larvae were collected, three larvae were found within a 15cm radius of collection Bzo6 and one larva was found within the same radius of collection Bzo8. Two additional larvae were collected randomly in leaf litter.

The larvae were collected in 50ml tubes to collect fresh faecal pellets. Fresh pellets and beetle larvae were stored in pure ethanol. Fresh fruiting bodies were stored in 50ml tubes and tissue cultures were made on the same day. Conidial cultures were made by streaking conidia from the pileus of the fruiting body on a petri dish containing the same medium. Cultures were stored at room temperature.

To detect potential faecal pellets and to study the function of the pseudorhiza of *Tephrocybe rancida*, we sampled mushrooms at one location on the property of the Nyenrode Business University in Breukelen, the Netherlands. We sampled in late October of 2015, 2017 and 2019. The rooting base was carefully excavated using a small gardening trowel. The pseudorhiza was traced into the soil but no clear connection to faecal pellets or any buried substrate could be found. Fruiting bodies were collected in 50ml tubes and were either dried or used to make tissue cultures. Single-spore isolates were made by attaching a pileus with petroleum jelly to the lid of a petri dish and spores were captured on MYA medium with streptomycin. When basidiospores had germinated a single colony was transferred to a new MYA plate.

Scoring of the morphological predisposition traits

The four morphological traits were scored after inspection of the specimens that were collected. In most cases the traits could unambiguously be detected from our collections. However, in some cases we were not able to score all four traits as we only had a culture and no mushroom or the other way around. In these cases, scoring was done combining our personal observations and those from literature.

Our single collection of *Tricholomella constricta* did not have a pseudorhiza; however, this trait is variable within this species[18], and therefore, we scored it as present . In the case of *Termitomyces* we made the general assumption that all species have a pseudorhiza although there may be a single case in which there may not always be a pseudorhiza, which is in *T. microcarpus*[58]. However, because some authors describe it as weakly rooting[26], it is probably still able to produce the pseudorhiza and we therefore scored it as present . The ambiguity of this trait in *T. microcarpus* is most likely a response to epigeous fruiting on expelled comb material.

We were not able to detect conidia in our culture of *Sphagnurus paluster;* however, conidial production has been reported for this species[59] and we therefore scored it as present .

The presence of clamp connections has been reported in the following genera or species: *Asterophora*[60], *Blastosporella*[25], *Calocybe cyanea*[61] and *Hypsizygus*[62], *Lyophyllum*[63], *Myochromella*[5], *Tricholomella*[18], *T. rancida*[5], *S. paluster*[64]. The absence of clamp connections is reported in the following genera: *Arthromyces*[25] and *Termitomyces*[32,58].

DNA isolation

For DNA isolation of the fungal samples a small piece (0.2-0.5g) of mycelium from a laboratory culture, the pileus of a dried herbarium specimen or the pileus of a specimen stored in ethanol was frozen in liquid nitrogen and disrupted in a 1.5ml Eppendorf tube with glass beads prior to DNA isolation. DNA isolation for all samples was performed using the cetyltrimethylammonium bromide (CTAB) as previously described[65]. DNA from beetle larvae was isolated from a leg part using the same protocol.

DNA from faecal pellets was isolated by using a Nucleospin Soil DNA extraction kit (Macherey-Nagel) following the manufacturer s instructions. For the fresh pellets that were directly obtained from living beetle larvae only a single pellet was used for DNA isolation, from the pellets that were part of the fungal substrate between five and 20 pellets were used depending on the size of the pellets and availability.

Beetle larvae identification

As the identification of beetles from the larval stage is very difficult and requires a field expert, we attempted molecular identification of the six larvae that we collected. This was done by obtaining a partial sequence of the mitochondrial cytochrome c oxidase (CO1) and using NCBI BLAST for identification. A partial sequence of the CO1 was amplified using the primer pair C1-J-2183 (Jerry)/ TL2-N-3014 (Pat)[66], using the following PCR program: denaturation at 94 C for 60 seconds, then five cycles consisting of 30 seconds denaturation at 94 C, 40 seconds annealing at 47 C and elongation for 60 seconds at 72 C, followed by 30 cycles consisting of 30 seconds denaturation at 94 C, 40 seconds annealing at 52 C and elongation for 60 seconds at 72 C, followed by a final extension step for 10 minutes at 72 C. After Sanger sequencing of amplified products, we were able to distinguish two different genotypes, of which four larvae with identical sequences were of genotype 1 (MW698941) and two larvae with identical sequences of genotype 2 (MW698942). We could not make a reliable identification for either genotype because searches against the Genbank database did not return a close match (97% similarity). The closest match for genotype 1 was a 85.75% match to a *Cryptodus* sp. sequence (KF801857), while genotype 2 had a 84.17% match to a *Pimelopus dubius dubius* sequence (EF487738). Both of these species belong to the subfamily Dynastinae (Rhinoceros beetles) within the Scarabaeidae.

Faecal pellet identification

Identification of the depositor of the faecal pellets of *B. zonata* or *Arthromyces* could shed light on the interaction between the fungus and the insect partner. As we were only able to obtain fresh pellets in ethanol for *B. zonata* we focused on these pellets. We attempted amplification of the mitochondrial cytochrome c oxidase (CO1) on DNA extracted from pellets using PCR with primer pairs, LCO1490/HCO2198[67] and Jerry/Pat[66]. We also attempted amplification of the internal transcribed spacer 1 (ITS1) marker using the Vogler primer pair[68] and the 16S marker using the Coleoptera specific primers Coleop_16Sc and Coleop_16Sd[69]. PCR was performed using protocols and conditions described for each primer pair specified in the original publications. We could not confidently observe amplification of host DNA as we obtained multiple different PCR products per reaction.

Environmental DNA barcoding

Our standard PCR protocol was insufficient to identify the depositor of the pellets so we opted for an environmental DNA barcoding approach. A recent study used environmental DNA barcoding approaches to determine dietary arthropod contents in faecal samples of insectivorous animals[70]. We used two different DNA barcodes: a 157bp target region of the CO1 using the primer pair ZBJ-ArtF1c/ZBJ-ArtR2c[71] and a 156bp target region of 16S using the primer pair Ins16S_1shortF/Ins16S_1shortR[72]. PCR was performed using protocols and conditions described for each primer pair specified in the original publications. A total of 19 PCR reactions were done using the protocol below on 13 different DNA samples; six

from a *B. zonata* pellet substrate, five from pellets not visibly colonized by *B. zonata* and two from fresh beetle pellets. Unique barcode adapters were used for each PCR reaction and all samples were pooled after PCR and sequenced using an Oxford nanopore MinION. To test whether the pellets were of beetle origin we used the sequences generated from the larvae that we collected as reference. We used Geneious 10.0.9 (www.geneious.com) to match reads (between 1,000 and 27,000 reads per sample) against the reference but were unable to find any significant matches (data not shown)[73]. Since we used two DNA samples from faecal pellets directly obtained from these larvae this suggests that our method may not be sensitive enough to pick up host DNA (from gut epithelial cells) from the faecal pellets.

Electron microscopy

Scanning electron microscopy on laboratory cultures of *B. zonata* (Bzo9), *A. matolae* (FLAS-F-62734) and *T. cryptogamus* (P5) was performed at the Wageningen Electron Microscopy Centre. To preserve the delicate conidiophores in *A. matolae* and the nodules in *T. cryptogamus* samples were frozen in liquid nitrogen prior to imaging (cryoSEM). The *A. matolae* culture that was used was grown for 14 days at 15°C on MYA agar. The *T. cryptogamus* culture was grown for 25 days at 25°C on MYA.

Library preparation and whole genome sequencing

Library preparation and whole-genome sequencing was performed by Novogene (Hong Kong) using the Illumina Hiseq 2500 platform. The paired-end reads that were generated were 150bp long and the insert size was 500bp.

Assembly and annotation

We assembled paired Illumina reads using SPAdes v.3.5.0 with default settings[74]. Short contigs smaller than 300bp or contigs with a coverage lower than 5x were removed from the assembly using a script. Presumed bacterial contigs were removed from the assembly using a script which matched contigs using BLAST against a reference library of 500 randomly selected bacterial genomes. Contigs with a BLAST hit with an expect value (E) of less than 1e-7 were removed from the assembly. We applied automatic annotations to each assembly using the funannotate pipeline (v.1.7.4). We ran funannotate -mask with default options for repeat masking, followed by funannotate -sort. We then ran funannotate -predict using a pretrained Augustus data set for *Termitomyces cryptogamus* T132 as reference[75]. We also ran funannotate -iprscan, funannotate -remote with antiSmash, and funannotate -annotate all with default settings[76].

Marker selection and phylogenetic analysis

We collected conserved orthologs using BUSCO with the basidiomycete reference gene set odb9 provided on the BUSCO website[77]. We then removed any sequence sets for which

we found fewer than 25 matches. Finally, we aligned the remaining sequence sets using MAFFT v.7.475[78] with the following parameters: --auto - maxiterate 1000 -adjustdirection

To remove poorly aligned regions, we used Gblocks v.0.91b[79] with the following input: -o -b5=h -t=DNA.

We then concatenated all trimmed alignments and ran a maximum likelihood phylogenetic analysis with IQ-TREE (version 2.1.2) with the following settings: -s -spp -o -bb 1000 -bsam GENESITE -m TESTMERGE --runs 100, with a partition for each BUSCO locus, and *Hypsizygus ulmarius* as outgroup. Using -bsam GENESITE reduces bootstrap inflation by resampling partitions first and then resampling sites within partitions. We ran 100 independent runs and all produced the same topology with minimal variation in likelihood estimates. To estimate phylogenetic conflicts between loci we compared species tree to individual locus trees estimated with IQ-TREE (version 2.1.2)[80]. We reconstructed the locus trees with the following parameters: -s -S. The gene and site concordance factors were then computed with IQTree using the command: -t -gcf -s --scf 100. In addition, we used the locus trees generated by IQ-TREE to perform a coalescent-based species tree reconstruction using ASTRAL[81]. We performed ASTRAL with default settings using the command: java -jar astral.5.6.3.jar -i -o 2>out.log.

CAZyme analysis

Predicted CAZymes were collected from the funannotate output for each assembly in our dataset. These predictions are made from the dbCAN2 database based on Hidden Markov Model predictions. To increase the confidence of the predictions, we then submitted the corresponding amino acid sequences to the dbCAN2 webserver, to obtain the predictions for this set of proteins using DIAMOND and Hotpep[82]. Only predicted CAZymes that were also identified using either the DIAMOND or Hotpep pipelines were used for the analysis.

CAFE analysis

To detect significant changes in CAZyme composition along our phylogenetic tree we ran CAFE[83] on our CAZyme dataset using default settings. The phylogenetic tree reconstructed using IQ-tree was made ultrametric using the package Ape in R and was used as input for the analysis[84].

Data S1 Raw CAZyme data showing the number of predicted CAZymes for each class for all species.

Data S2 Sample table with metadata and NCBI Genbank accession numbers for each species.

References

1. Bignell, D.E. (2000). Introduction to Symbiosis. In Termites: Evolution, Sociality, Symbioses, Ecology, T. Abe, D.E. Bignell, and M. Higashi, eds. (Springer Netherlands), pp. 189-208. 10.1007/978-94-017-3223-9_9.
2. Chouvenc, T., Sobotnik, J., Engel, M.S., and Bourguignon, T. (2021). Termite evolution: mutualistic associations, key innovations, and the rise of Termitidae. Cellular and Molecular Life Sciences *78*, 2749-2769. 10.1007/s00018-020-03728-z.
3. Nobre, T., and Aanen, D.K. (2012). Fungiculture or Termite Husbandry? The Ruminant Hypothesis. Insects *3*, 307-323. 10.3390/insects3010307.
4. Aanen, D.K., Eggleton, P., Rouland-Lefevre, C., Guldberg-Froslev, T., Rosendahl, S., and Boomsma, J.J. (2002). The evolution of fungus-growing termites and their mutualistic fungal symbionts. Proceedings of the National Academy of Sciences of the United States of America *99*, 14887-14892. 10.1073/pnas.222313099.
5. Hofstetter, V., Redhead, S.A., Kauff, F., Moncalvo, J.M., Matheny, P.B., and Vilgalys, R. (2014). Taxonomic revision and examination of ecological transitions of the Lyophyllaceae (Basidiomycota, Agaricales) based on a multigene phylogeny. Cryptogamie Mycologie *35*, 399-425. 10.7872/crym.v35.iss4.2014.399.
6. Hofstetter, V., Clemencon, H., Vilgalys, R., and Moncalvo, J.M. (2002). Phylogenetic analyses of the Lyophylleae (Agaricales, Basidiomycota) based on nuclear and mitochondrial rDNA sequences. Mycological Research *106*, 1043-1059. 10.1017/S095375620200641x.
7. Matheny, P.B., Curtis, J.M., Hofstetter, V., Aime, M.C., Moncalvo, J.M., Ge, Z.W., Slot, J.C., Ammirati, J.F., Baroni, T.J., Bougher, N.L., et al. (2006). Major clades of Agaricales: a multilocus phylogenetic overview. Mycologia *98*, 982-995. 10.3852/mycologia.98.6.982.
8. Bellanger, J.M., Moreau, P.A., Corriol, G., Bidaud, A., Chalange, R., Dudova, Z., and Richard, F. (2015). Plunging hands into the mushroom jar: a phylogenetic framework for Lyophyllaceae (Agaricales, Basidiomycota). Genetica *143*, 169-194. 10.1007/s10709-015-9823-8.
9. Kubatko, L.S., and Degnan, J.H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology *56*, 17-24. 10.1080/10635150601146041.
10. An , C., Larget, B., Baum, D.A., Smith, S.D., and Rokas, A. (2007). Bayesian estimation of concordance among gene trees. Molecular Biology and Evolution *24*, 412-426. 10.1093/molbev/msl170.
11. Minh, B.Q., Hahn, M.W., and Lanfear, R. (2020). New Methods to Calculate Concordance Factors for Phylogenomic Datasets. Molecular Biology and Evolution *37*, 2727-2733. 10.1093/molbev/msaa106.
12. Harrington, T.C. (1987). New Combinations in Ophiostoma of Ceratocystis Species with Leptographium Anamorphs. Mycotaxon *28*, 39-43.
13. Harrington, T.C. (2005). Ecology and evolution of mycophagous bark beetles and their fungal partners. In Insect-Fungal Associations: Ecology and Evolution, M.B. F.E. Vega, ed. (Oxford Univeristy Press), pp. 257-291.

14. Mueller, U.G., Gerardo, N.M., Aanen, D.K., Six, D.L., and Schultz, T.R. (2005). The evolution of agriculture in insects. Annu Rev Ecol Evol S *36*, 563-595. 10.1146/annurev.ecolsys.36.102003.152626.
15. De Duve, C. (2005). Singularities: landmarks on the pathways of life (Cambridge University Press).
16. Burnetti, A., and Ratcliff, W.C. (2020). The Origin of Phototrophy Reveals the Importance of Priority Effects for Evolutionary Innovation. Preprints.
17. Buller, A.H.R. (1924). Agarics which are parasitic on other agarics (Longmans Green and Co.).
18. Kalamees, K. (1992). Tricholomella, a New Genus, with the Distribution Data of Tricholomella-Constrictum, Comb-Nov in East Europe and Asia. Persoonia *14*, 445-447.
19. Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R.A., Henrissat, B., Martinez, A.T., Otillar, R., Spatafora, J.W., Yadav, J.S., et al. (2012). The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science *336*, 1715-1719. 10.1126/science.1221748.
20. Poulsen, M., Hu, H.F., Li, C., Chen, Z.S., Xu, L.H., Otani, S., Nygaard, S., Nobre, T., Klaubauf, S., Schindler, P.M., et al. (2014). Complementary symbiont contributions to plant decomposition in a fungus-farming termite. Proceedings of the National Academy of Sciences of the United States of America *111*, 14500-14505. 10.1073/pnas.1319718111.
21. Revell, L.J. (2009). Size-correction and principal components for interspecific comparative studies. Evolution *63*, 3258-3268. 10.1111/j.1558-5646.2009.00804.x.
22. K es, U., Badalyan, S.M., Gie ler, A., and D rnte, B. (2016). 12 Asexual Sporulation in Agaricomycetes. In The Mycota I, 3rd edition. Growth, Differentiation and Sexuality, W. J., ed. (Springer), pp. 269-328.
23. Botha, W.J., and Eicker, A. (1991). Cultural studies on the genus Termitomyces in South Africa. I. Macro-and microscopic characters of basidiome context cultures. Mycological Research *95*, 435-443.
24. Hinze, B., Crailsheim, K., and Leuthold, R.H. (2002). Polyethism in food processing and social organisation in the nest of Macrotermes bellicosus (Isoptera, Termitidae). Insectes Sociaux *49*, 31-37. 10.1007/s00040-002-8275-1.
25. Baroni, T.J., Franco-Molano, A.E., Lodge, D.J., Lindner, D.L., Horak, E., and Hofstetter, V. (2007). Arthromyces and Blastosporella, two new genera of conidia-producing lyophylloid agarics (Agaricales, Basidiomycota) from the neotropics. Mycological research *111*, 572-580. 10.1016/j.mycres.2007.03.007.
26. Horak, E. (1968). Synopsis Generum Agaricalium (Die Gattungstypen der Agaricales). Beitrage zur Kryptogamenflora der Schweiz *13*.
27. Buller, A.H.R. (1934). Researches on fungi VI (Longmans, Green & Company).
28. Campbell, A.H. (1938). Contribution to the biology of Collybia radicata (Relh.) Berk. Transactions of the British Mycological Society *22*, 151-159. 10.1080/00275514.1991.12025978.
29. Redhead, S.A., and Malloch, D.W. (1986). The genus Phaeocollybia (Agaricales) in eastern Canada and its biological status. Canadian Journal of Botany *64*, 1249-1254. 10.1139/b86-172.

30. Sagara, N., Hongo, T., Murakami, Y., Hashimoto, T., Nagamasu, H., Fukiharu, T., and Asakawa, Y. (2000). Hebeloma radicosoides sp nov., an agaric belonging to the chemoecological group ammonia fungi. Mycological Research *104*, 1017-1024. 10.1017/S0953756299002439.
31. Sagara, N., Senn-Irlet, B., and Marstad, P. (2006). Establishment of the case of Hebeloma radicosum growth on the latrine of the wood mouse. Mycoscience *47*, 263-268.
32. Nobre, T., Koopmanschap, B., Baars, J.J.P., Sonnenberg, A.S.M., and Aanen, D.K. (2014). The scope for nuclear selection within Termitomyces fungi associated with fungus-growing termites is limited. Bmc Evol Biol *14*, 1-12. 10.1186/1471-2148-14-121.
33. Kooij, P.W., Aanen, D.K., Schi tt, M., and Boomsma, J.J. (2015). Evolutionarily advanced ant farmers rear polyploid fungal crops. Journal of evolutionary biology *28*, 1911-1924. 10.1111/jeb.12718.
34. Kuhn, G., Hijri, M., and Sanders, I.R. (2001). Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. Nature *414*, 745-748. 10.1038/414745a.
35. Bendich, A.J. (1987). Why do chloroplasts and mitochondria contain so many copies of their genome? BioEssays *6*, 279-282.
36. Bretagnolle, F., and Thompson, J.D. (1995). Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants. New Phytologist *129*, 1-22. DOI 10.1111/j.1469-8137.1995.tb03005.x.
37. De Fine Licht, H.H., Andersen, A., and Aanen, D.K. (2005). Termitomyces sp. associated with the termite Macrotermes natalensis has a heterothallic mating system and multinucleate cells. Mycological Research *109*, 314-318. 10.1017/s0953756204001844.
38. Zeder, M.A. (2012). Pathways to animal domestication. Biodiversity in agriculture: domestication, evolution, and sustainability, 227-259.
39. Zeder, M.A. (2015). Core questions in domestication research. Proceedings of the National Academy of Sciences *112*, 3191-3198. 10.1073/pnas.1501711112.
40. Viana, A.B., C rtes, M.O., Cornelissen, T.G., and Neves, F.D. (2018). Interactions between wood-inhabiting fungi and termites: a meta-analytical review. Arthropod-Plant Interactions *12*, 229-235. 10.1007/s11829-017-9570-0.
41. Waller, D.A., Lafage, J.P., Gilbertson, R.L., and Blackwell, M. (1987). Wood-Decay Fungi Associated with Subterranean Termites (Rhinotermitidae) in Louisiana. Proceedings of the Entomological Society of Washington *89*, 417-424.
42. Chiu, C.I., Ou, J.H., Chen, C.Y., and Li, H.F. (2019). Fungal nutrition allocation enhances mutualism with fungus-growing termite. Fungal Ecology *41*, 92-100. 10.1016/j.funeco.2019.04.001.
43. Batra, L.R., and Batra, S.W.T. (1979). Termite-fungus mutualism. held in University of South Florida, Tampa, Florida, L.R. Batra, ed. (Wiley, New York), pp. 117-163.
44. Rouland-Lef vre, C. (2000). Symbiosis with fungi. In Termites: evolution, sociality, symbioses, ecology, (Springer), pp. 289-306.
45. Brugerolle, G., and Radek, R. (2006). Symbiotic protozoa of termites. In Intestinal microorganisms of termites and other invertebrates, (Springer), pp. 243-269.

46. Darlington, J.P.E.C. (1994). Nutrition and evolution in fungus-growing termites. Nourishment and evolution in insect societies, 105-130.
47. Eggleton, P. (2006). The termite gut habitat: its evolution and co-evolution. In Intestinal microorganisms of termites and other invertebrates, H. K nig, and A. Varma, eds. (Springer), pp. 373-404.
48. Inward, D.J., Vogler, A.P., and Eggleton, P. (2007). A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. Mol Phylogenet Evol *44*, 953-967. 10.1016/j.ympev.2007.05.014.
49. Garnier-Sillam, E., Toutain, F., Villemin, G., and Renoux, J. (1989). tudes Pr liminaires des Meules Originales du Termite XylophageSphaerotermes sphaerothorax (Sjostedt). Insectes sociaux *36*, 293-312.
50. Bucek, A., Sobotnik, J., He, S., Shi, M., McMahon, D.P., Holmes, E.C., Roisin, Y., Lo, N., and Bourguignon, T. (2019). Evolution of Termite Symbiosis Informed by Transcriptome-Based Phylogenies. Current Biology *29*, 3728-3734 e3724. 10.1016/j.cub.2019.08.076.
51. Sigoillot, J., Berrin, J., Bey, M., Lesage-Meessen, L., Levasseur, A., Lomascolo, A., Record, E., and Uzan-Boukhris, E. (2012). Fungal strategies for lignin degradation. In Advances in botanical research, (Elsevier), pp. 263-308.
52. Redhead, S.A., Ammirati, J.F., Walker, G.R., Norvell, L.L., and Puccio, M.B. (1994). Squamanita contortipes, the Rosetta Stone of a mycoparasitic agaric genus. Canadian Journal of Botany *72*, 1812-1824. 10.1139/b94-223.
53. Smith, M.E., Henkel, T.W., and Rollins, J.A. (2015). How many fungi make sclerotia? Fungal Ecology *13*, 211-220.
54. Diamond, J.M., and Ordunio, D. (1999). Guns, germs, and steel (Books on Tape).
55. Purugganan, M.D., and Fuller, D.Q. (2009). The nature of selection during plant domestication. Nature *457*, 843-848. 10.1038/nature07895.
56. Diamond, J. (2002). Evolution, consequences and future of plant and animal domestication. Nature *418*, 700-707. 10.1038/nature01019.
57. Leuthold, R.H., Badertscher, S., and Imboden, H. (1989). The inoculation of newly formed fungus comb with Termitomyces in Macrotermes colonies (Isoptera, Macrotermitinae). Insectes Sociaux *36*, 328-338.
58. Heim, R. (1977). Termites et champignons; les champignons termitophiles d'Afrique noire et d'Asie meridionale (Editions Boub e).
59. Walther, G., Garnica, S., and Weiss, M. (2005). The systematic relevance of conidiogenesis modes in the gilled Agaricales. Mycological research *109*, 525-544. 10.1017/s0953756205002868.
60. Thompson, G.E. (1936). Nyctalis parasitica and N. asterophora in culture. Mycologia *28*, 222-227.
61. Baroni, T.J., Legon, N.W., Vilgalys, R., and Lodge, D.J. (1999). Calocybe cyanea - a rare and beautiful agaric is discovered in Puerto Rico. Mycologist *13*, 7.
62. Singer, R. (1947). New genera of fungi. III. Mycologia *39*, 77-89.

63. Moncalvo, J.M., Rehner, S.A., and Vilgalys, R. (1993). Systematics of Lyophyllum Section Difformia Based on Evidence from Culture Studies and Ribosomal DNA-Sequences. Mycologia *85*, 788-794. 10.2307/3760610.
64. Redhead, S.A. (1981). Parasitism of Bryophytes by Agarics. Canadian Journal of Botany *59*, 63-67. DOI 10.1139/b81-011.
65. Nieuwenhuis, M., van de Peppel, L.J.J., Bakker, F.T., Zwaan, B.J., and Aanen, D.K. (2019). Enrichment of G4DNA and a Large Inverted Repeat Coincide in the Mitochondrial Genomes of Termitomyces. Genome Biology and Evolution *11*, 1857-1869. 10.1093/gbe/evz122.
66. Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., and Flook, P. (1994). Evolution, Weighting, and Phylogenetic Utility of Mitochondrial Gene-Sequences and a Compilation of Conserved Polymerase Chain-Reaction Primers. Annals of the Entomological Society of America *87*, 651-701. 10.1093/aesa/87.6.651.
67. Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology *3*, 294-299.
68. Vogler, A.P., and DeSalle, R. (1994). Evolution and phylogenetic information content of the ITS-1 region in the tiger beetle Cicindela dorsalis. Molecular Biology and Evolution *11*, 393-405. 10.1093/oxfordjournals.molbev.a040121.
69. Epp, L.S., Boessenkool, S., Bellemain, E.P., Haile, J., Esposito, A., Riaz, T., Erseus, C., Gusarov, V.I., Edwards, M.E., Johnsen, A., et al. (2012). New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. Mol Ecol *21*, 1821-1833. 10.1111/j.1365-294X.2012.05537.x.
70. Jusino, M.A., Banik, M.T., Palmer, J.M., Wray, A.K., Xiao, L., Pelton, E., Barber, J.R., Kawahara, A.Y., Gratton, C., Peery, M.Z., and Lindner, D.L. (2019). An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. Molecular ecology resources *19*, 176-190. 10.1111/1755-0998.12951.
71. Zeale, M.R., Butlin, R.K., Barker, G.L., Lees, D.C., and Jones, G. (2011). Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. Molecular ecology resources *11*, 236-244. 10.1111/j.1755-0998.2010.02920.x.
72. Clarke, L.J., Soubrier, J., Weyrich, L.S., and Cooper, A. (2014). Environmental metabarcodes for insects: in silico PCR reveals potential for taxonomic bias. Molecular ecology resources *14*, 1160-1170. 10.1111/1755-0998.12265.
73. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647-1649. 10.1093/bioinformatics/bts199.
74. Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Prjibelsky, A., Pyshkin, A., Sirotkin, A., Sirotkin, Y., et al. (2013). Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. held in Berlin, Heidelberg, (Springer Berlin Heidelberg), pp. 158-170.

75. Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Research *33*, W465-W467. 10.1093/nar/gki458.
76. Palmer, J.S., J. (2019). nextgenusfs/funannotate: funannotate v1.5.3 (Version 1.5.3). Zenodo. 10.5281/zenodo.2604804.
77. Seppey, M., Manni, M., and Zdobnov, E.M. (2019). BUSCO: assessing genome assembly and annotation completeness. In Gene Prediction, (Springer), pp. 227-245.
78. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution *30*, 772-780. 10.1093/molbev/mst010.
79. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution *17*, 540-552. 10.1093/oxfordjournals.molbev.a026334.
80. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution *37*, 1530-1534.
81. Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., and Warnow, T. (2014). ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics *30*, i541-548. 10.1093/bioinformatics/btu462.
82. Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P.K., Xu, Y., and Yin, Y. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic acids research *46*, W95-W101. 10.1093/nar/gky418.
83. De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics *22*, 1269-1271. 10.1093/bioinformatics/btl097.
84. Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526-528. 10.1093/bioinformatics/bty633.

Supplementary Data

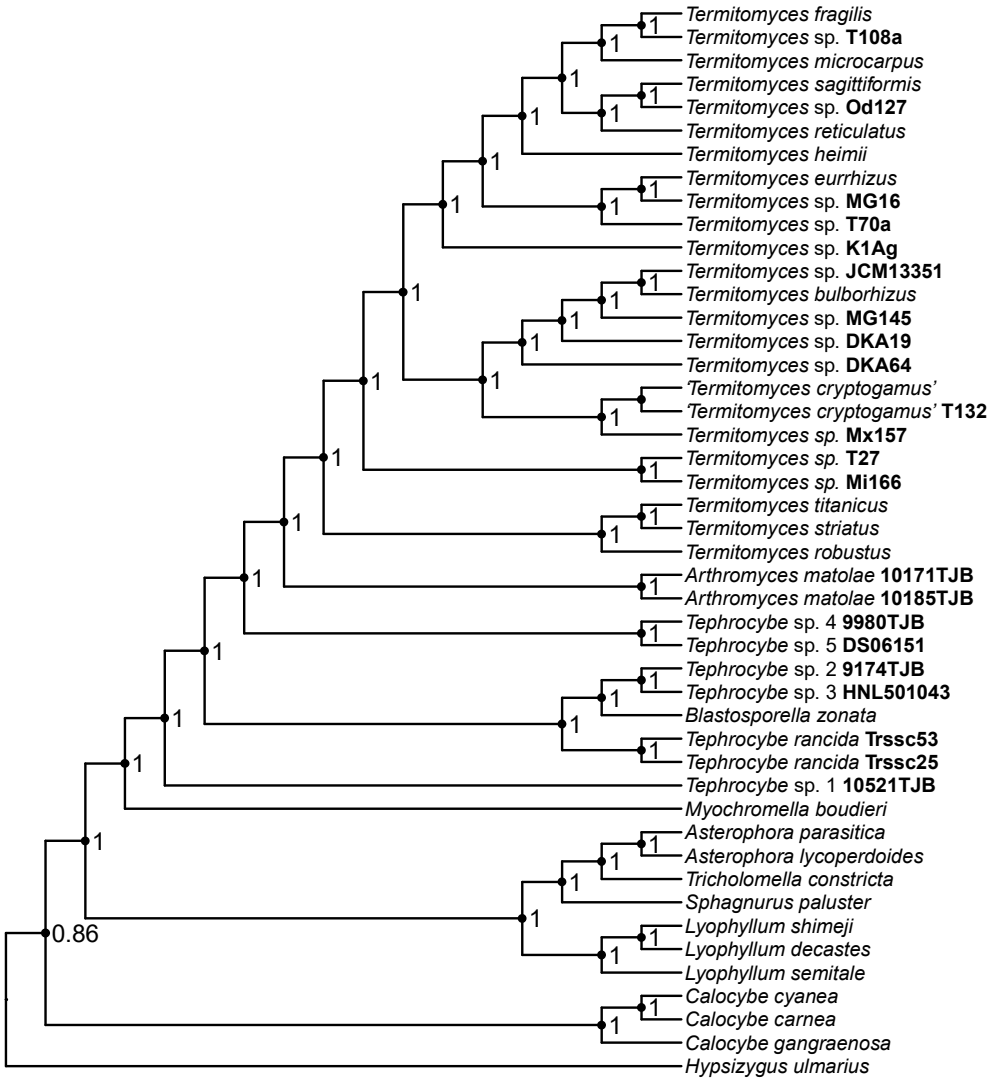


Figure S1: Phylogenetic tree resulting from coalescent-based ASTRAL analysis based on the same dataset used for Figure 1.

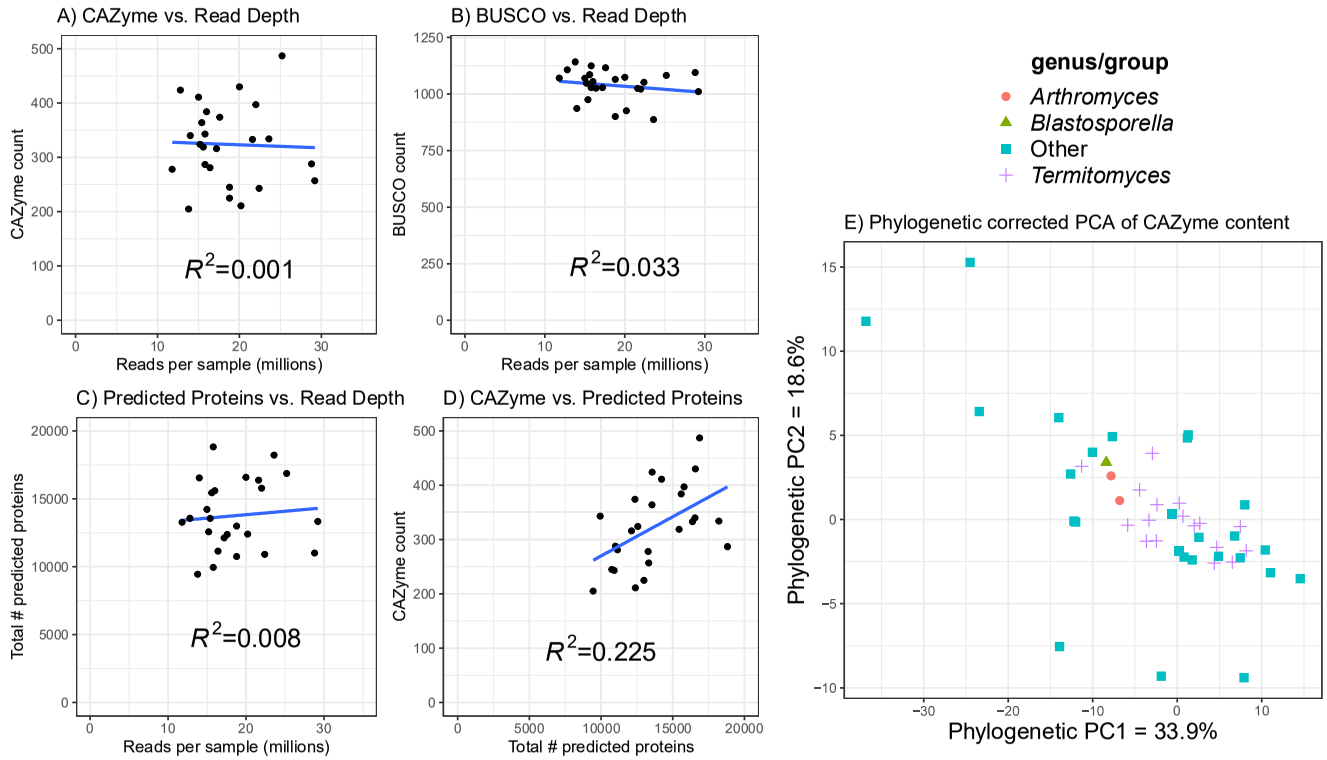


Figure S2: Correlation between read depth and predicted CAZymes and PCA of CAZyme content corrected for phylogeny. Related to Figure 1. Plots showing no significant correlation between the number of reads used in the assemblies and the content of CAZymes (A), BUSCO genes (B) and predicted proteins (C). Plot D shows the relationship between the number of predicted proteins and CAZyme content. Plot E shows a principal component analysis on CAZyme content corrected for phylogeny. Species of *Arthromyces* are represented by red circles, *Blastosporella zonata* is represented by a green triangle, species of *Termitomyces* by purple plus signs and all other lyophylloid taxa are represented by blue squares.

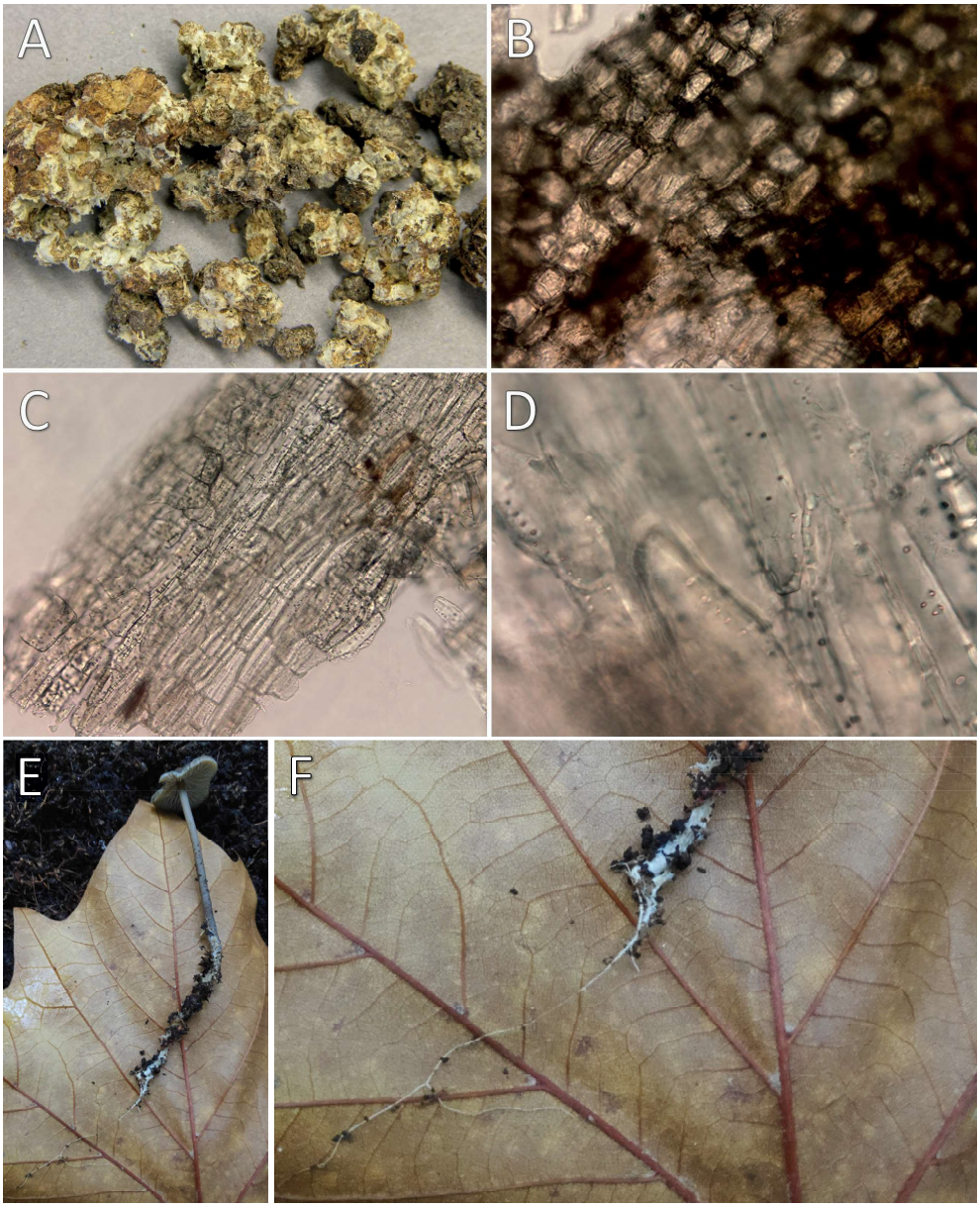


Figure S3: Insect faecal substrate of *Arthromyces matolae* and mushroom of *Tephrocybe rancida*. Related to Figure 2. Macroscopic picture of faecal substrate (A). Microscopic picture showing cork-like structure (B), vessel elements (C) and (bordered) pits in the cell walls (D). These elements are characteristic for plant xylem indicating a woody origin of the substrate. Mushroom of *Tephrocybe rancida* with tapering pseudorhiza (E). After careful excavation of the pseudorhiza we could not find any attachment to a buried substrate as the pseudorhiza tapered into a thin thread which would eventually break (F).

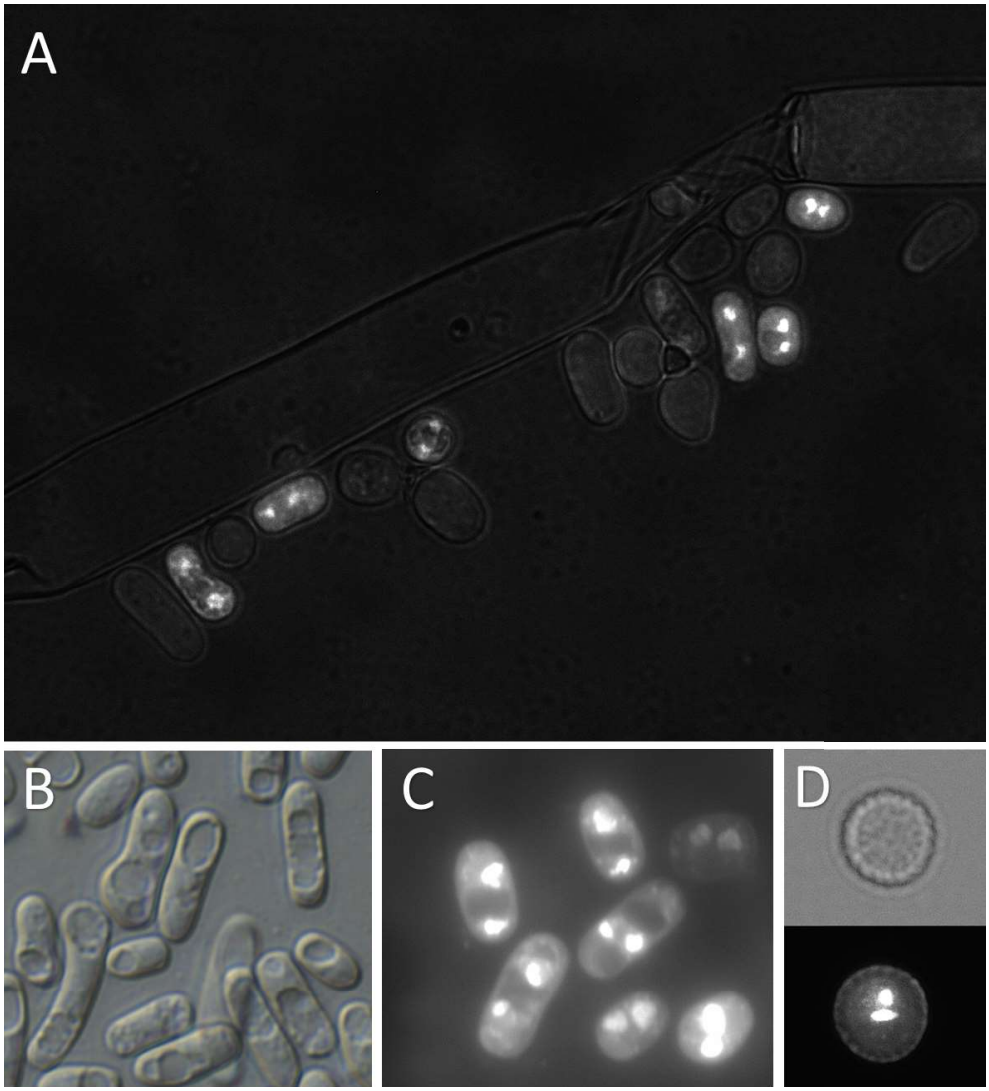


Figure S4: Production of asexual spores (conidia) in laboratory cultures. Related to Figure 3. Dapi staining of the arthroconidia produced in a culture of *Arthromyces matolae* (FLAS-F-62734) showing two nuclei per conidium (A). Dikaryotic arthroconidia (B and C) and dikaryotic blastoconidia (D) produced in a culture of *Blastosporella zonata* (Bzo9).

CHAPTER 4

4

Transfer of mitochondrial tRNA genes to linear plasmids in fungi facilitates loss of such genes from mitochondrial DNA

Mathijs Nieuwenhuis, Jeroen Groeneveld, Duur K. Aanen

Abstract

Abstract

Fungal and plant mitochondria are known to exchange DNA with retroviral plasmids. Transfer of plasmid DNA to the organellar genome is best known and occurs through wholesale insertion of the plasmid. Less well known is the transfer of organellar DNA to the plasmid, presumably by incorporation of organellar RNA, in particular encoding tRNA products, during the replication of the plasmid. In this paper, we report at least six independent transfers of complete tRNA genes to fungal plasmids. Furthermore, we discovered two independent cases of loss of a tRNA gene from a fungal mitochondrial genome following transfer of such a gene to a linear mitochondrial plasmid. We propose that loss of a tRNA gene from mtDNA following its transfer to a plasmid creates a mutualistic dependency of the host mtDNA on the plasmid. We find that tRNA genes transferred to plasmids encode codons that occur at the lowest frequency in the host mitochondrial genomes. We discuss the potential consequences of mtDNA transfer to plasmids for both the host mtDNA and the plasmid.

Introduction

Mitochondria of plants and fungi are frequently hosts to retroviral plasmids. These plasmids can be divided into two classes based on their structure: circular and linear, with linear plasmids further subdivided into invertron and non-invertron like types (Griffiths 1995). Circular plasmids appear to use a rudimentary process of reverse transcription to replicate themselves (Galligan & Kennell 2007; Kennell et al. 1994). Invertron-like plasmids encode a DNA and RNA polymerase in inverted orientation to each other, and feature terminal inverted repeats bound covalently to proteins at the 5′ end. Invertron plasmids are thought to replicate in similar fashion to adenoviruses and certain bacteriophages, which share a similar structure (Griffiths 1995; Kempken et al. 1989), and their replication involves a protein primer. For invertron plasmids, the proteins bound to the terminal repeats probably play a role in plasmid replication.

Both circular and linear plasmids can insert themselves wholly or partially in their host mitochondrial DNA (mtDNA). In some fungi, the insertions of plasmids in mitochondrial DNA (mtDNA) are linked to phenotypic effects, most notably senescence in *Neurospora* and *Podospora* (Griffiths et al. 1990; Tudzynski et al. 1980; van Diepeningen et al. 2008; Maas et al. 2004, 2005). However, in most fungi the presence and occasional mitochondrial insertion of these plasmids is assumed to be inconsequential for the fungal phenotype. Plasmids can infect new hosts through horizontal transfer (Kempken 1995) but are also occasionally removed from their hosts during vertical transfer (Van Der Gaag et al. 1998). To persist in a host population the spread of plasmids via horizontal transmission must therefore offset the loss through vertical transmission.

A limited number of studies report transfer of mtDNA sequences to a plasmid. In *Neurospora*, partial mtDNA sequences were detected in circular plasmids of the *Varkud* and *Mauriceville* types, in addition to inserts of other, linear plasmid types (R A Akins et al. 1989; Mohr et al. 2000). In the circular plasmid *Mauriceville*, Chiang and Lambowitz (Chiang & Lambowitz 1997) found that reverse transcriptase activity was initiated without the use of a primer (as is common in reverse transcriptases) at the 3′ end of a tRNA-like structure on the plasmid. The plasmid reverse transcriptase could also begin at the 3′ end of a tRNA gene, resulting in cDNA synthesis of that tRNA gene. Accidental inclusion of a mitochondrial tRNA transcript could result from the reverse transcriptase mistaking the tRNA for the plasmid-encoded tRNA-like structure. This would explain the origin of the tRNA genes found in these circular plasmids. Kempken (Kempken 1989) reported a tRNA-like structure on a linear plasmid found in *Ascobolus immersus*. If linear plasmids employ tRNA-like structures as initiation sites for replication, in similar fashion to circular plasmids, occasional transfer of mtDNA to linear plasmids is expected.

To our knowledge only one study has reported transfer of a mitochondrial tRNA gene to a linear plasmid. In maize, a mitochondrial tRNA gene was found in a non-invertron type linear plasmid (Leon et al. 1989). In the wild maize relative *Teosinte*, this plasmid-encoded tRNA gene has coincided with a loss of that tRNA gene in the mtDNA. This suggests the plasmid has adopted the tRNA function, assuming the tRNA is still required for the mitochondrion to function. If the organism becomes dependent on the plasmid for the production of this tRNA, it would present an interesting and rapidly evolving case of genetic addiction (Kobayashi 2004).

Addiction to a linear mitochondrial plasmid may be an important mechanism for their continued persistence. By acquiring a function that is vital or at least beneficial to the host, a plasmid can increase its chances of vertical transmission. Furthermore, transfer of mitochondrial DNA to linear plasmids may provide insight into their replication process. We studied the DNA exchange between fungal mitochondria and their associated linear plasmids, focussing in particular on transfer of mtDNA to plasmids. We focussed on fungi of the Lyophyllaceae family as a whole-genome sequencing data set generated for a phylogenetic study (Van de Peppel, unpublished) was available to screen for plasmids, and several complete mitochondrial genomes for this family are available to check for mtDNA transfer (Nieuwenhuis et al. 2019). In addition, Nieuwenhuis et al. already reported four linear mitochondrial plasmids and numerous mitochondrial plasmid inserts in their study, suggesting species in this family are often associated with such plasmids.

Methods

Samples and sequencing

We obtained sequencing data of 44 Lyophyllaceae strains, most of which represent distinct species, from SRA to screen for linear mitochondrial plasmids (Table 1). Furthermore, we sequenced whole genomic DNA of 10 additional *Termitomyces* strains. To estimate intraspecific variation of plasmids, one *Termitomyces* species was represented by multiple strains in our data set (*Termitomyces cryptogamus* J132/T132/T99/T8/T28/T61/T60a). Data consisted of paired-end Illumina reads (read length: 150bp, insert size: 500bp). We acquired 18 complete sequences from GenBank of known linear mitochondrial plasmids occurring in fungi (Table 1). We also included a plasmid sequence discovered in a sample of *Trametes versicolor* in our lab. All plasmid sequences we discovered in this study have been submitted to GenBank under the accession numbers MW874118- MW874172 (Table 1). Since the majority of plasmids used in this study were found in the Lyophyllaceae, we used twelve mitochondrial genomes of Lyophyllaceae (mostly *Termitomyces*) to analyse DNA transfer between mtDNA and plasmids (Nieuwenhuis et al. 2019).

Table 1: Linear mitochondrial plasmid sequences used in this study.

Plasmid	GenBank accession number	Host species	SRA accession number	Source
Podospora anserina pAL2-1	X60707.1	*Podospora anserina*	n.a.	(Hermanns & Osiewacz 1992)
Blumeria graminis pBgh	AY189817.1	*Blumeria graminis f.sp. hordei*	n.a.	(Giese et al. 2003)
Claviceps purpurea pClK1	X15648.1	*Claviceps purpurea*	n.a.	(Oeser & Tudzynski 1989)
Fusarium proliferatum pFP1	EF622512.1	*Fusarium proliferatum*	n.a.	(Láday et al. 2008)
Moniliophthora roreri pMR1	HQ259116.1	*Moniliophthora roreri*	n.a.	(Costa et al. 2012)
Lentinula edodes pLE	AB697990.1	*Lentinula edodes*	n.a.	(Yang et al. 2017)
Moniliophthora roreri pMR3	HQ259118.1	*Moniliophthora roreri*	n.a.	(Costa et al. 2012)
Pleurotus ostreatus mlp2	AF355103.1	*Pleurotus ostreatus*	n.a.	(Kim et al. 2000)
Flammulina velutipes pFV	AB028633.1	*Flammulina velutipes*	n.a.	(Nakai et al. 2000)
Gelasinospora Gel-kalDNA	L40494.1	*Gelasinospora sp. G114*	n.a.	(Yuewang et al. 1996)
Neurospora intermedia kalilo	X52106.1	*Neurospora intermedia*	n.a.	(Shiu-Shing Chan et al. 1991)
Neurospora crassa maranhar	X55361.1	*Neurospora crassa*	n.a.	(Court & Bertrand 1992)
Neurospora intermedia Harbin-3	AF133505.1	*Neurospora intermedia*	n.a.	(Xu et al. 1999)
Flammulina velutipes pFV2	AB028634.1	*Flammulina velutipes*	n.a.	(Nakai et al. 2000)
Hebeloma circinans pHC2	Y11504.1	*Hebeloma circinans*	n.a.	(Bai et al. 1997)
Moniliophthora roreri pMR2	HQ259117.1	*Moniliophthora roreri*	n.a.	(Costa et al. 2012)
Pleurotus ostreatus mlp1	AF126285.1	*Pleurotus ostreatus*	n.a.	Unpublished
Agaricus bitorquis pAB	X63075.1	*Agaricus bitorquis*	n.a.	(Robison et al. 1991)
Arthromyces claviformis pAC	MW874131	*Arthromyces claviformis*	SRX10708852	
Asterophora parasitica pAP	MW874136	*Asterophora parasitica*	SRX4910414	
Calocybe cyanea pCC	MW874161	*Calocybe cyanea*	SRX10337368	
Calocybe cyanea pCC2	MW874124	*Calocybe cyanea*	SRX10337368	
Calocybe gambosa pCG	MW874162	*Calocybe gambosa*	n.a.	
Calocybe gangraenosa pCGr	MW874123	*Calocybe gangraenosa*	SRX10337362	
Lyophyllum semitale pLSe	MW874135	*Lyophyllum semitale*	SRX10337363	
Lyophyllum shimeji pLS	MW874160	*Lyophyllum shimeji*		
Tephrocybe p9980TJB	MW874118	*Tephrocybe* sp. p9980TJB	SRX10337360	

Termitomyces bulborhizus pTB	MW874132	*Termitomyces bulborhizus*	SRX10337358
Termitomyces bulborhizus pTB	MW874159	*Termitomyces bulborhizus*	SRX10337358
Termitomyces DKA64 pDKA64 1	MW874172	*Termitomyces* sp. DKA64	SRX10337366
Termitomyces DKA64 pDKA64 2	MW874153	*Termitomyces* sp. DKA64	SRX10337366
Termitomyces DKA64 pDKA64 3	MW874155	*Termitomyces* sp. DKA64	SRX10337366
Termitomyces DKA64 pDKA64 4	MW874122	*Termitomyces* sp. DKA64	SRX10337366
Termitomyces J132 pJ132	MW874144	*Termitomyces* sp. J132	SRX255527
Termitomyces K1Aa pK1Aa 1	MW874149	*Termitomyces* sp. K1Aa	SRX10313007
Termitomyces K1Aa pK1Aa 2	MW874169	*Termitomyces* sp. K1Aa	SRX10313007
Termitomyces K1Aa pK1Aa 3	MW874137	*Termitomyces* sp. K1Aa	SRX10313007
Termitomyces K1Ac pK1Ac mitochondrion	MW874150	*Termitomyces* sp. K1Ac	SRX10337355
Termitomyces K1Ag pK1Ag 1	MW874133	*Termitomyces* sp. K1Ag	SRX10337354
Termitomyces K1Ag pK1Ag 2	MW874138	*Termitomyces* sp. K1Ag	SRX10337354
Termitomyces K2P1 pK2P1	MW874139	*Termitomyces* sp. K2P1	SRX10767240
Termitomyces Mi1657 pTMi1657	MW874127	*Termitomyces* sp. Mi1657	SRX10767235
Termitomyces pTDKA19	MW874119	*Termitomyces* sp. DKA19	SRX4910405
Termitomyces T108 pT108 1	MW874151	*Termitomyces* sp. T108	SRX10337365
Termitomyces T123 pT123 1	MW874120	*Termitomyces* sp. T123	SRX4910411
Termitomyces T123 pT123 2	MW874130	*Termitomyces sp.* T123	SRX4910411
Termitomyces T123 pT123 3	MW874154	*Termitomyces* sp. T123	SRX4910411
Termitomyces T123 pT123 4	MW874158	*Termitomyces* sp. T123	SRX4910411
Termitomyces T123 pT123 5	MW874140	*Termitomyces* sp. T123	SRX4910411
Termitomyces T13 pT13	MW874128	*Termitomyces* sp. T13	SRX4910413
Termitomyces T132 pT132 1	MW874146	*Termitomyces* sp. T132	SRX4910412
Termitomyces T132 pT132 2	MW874165	*Termitomyces* sp. T132	SRX4910412
Termitomyces T28 pT28 1	MW874163	*Termitomyces* sp. T28	SRX10474844
Termitomyces T28 pT28 3	MW874141	*Termitomyces* sp. T28	SRX10474844
Termitomyces T32 pT32	MW874134	*Termitomyces* sp. T32	SRX4910409
Termitomyces T40b pT40b 1	MW874125	*Termitomyces* sp. T40b	SRX10767241
Termitomyces T50a pT50a	MW874126	*Termitomyces* sp. T50a	SRX10767238
Termitomyces T60a pT60a 1	MW874143	*Termitomyces* sp. T60a	SRX10474845
Termitomyces T60a pT60a 2	MW874166	*Termitomyces* sp. T60a	SRX10474845

Termitomyces T61 pT61 1	MW874142	*Termitomyces* sp. T61	SRX10474846
Termitomyces T61 pT61 2	MW874164	*Termitomyces* sp. T61	SRX10474846
Termitomyces T70a pT70a 1	MW874171	*Termitomyces* sp. T70a	SRX10337364
Termitomyces T70a pT70a 2	MW874129	*Termitomyces* sp. T70a	SRX10337364
Termitomyces T70a pT70a 3	MW874152	*Termitomyces* sp. T70a	SRX10337364
Termitomyces T73sscA pT73sscA	MW874148	*Termitomyces* sp. T73sscA	SRX10313000
Termitomyces T8 pT8 1	MW874157	*Termitomyces* sp. T8	SRX10474843
Termitomyces T8 pT8 2	MW874168	*Termitomyces* sp. T8	SRX10474843
Termitomyces T8 pT8 3	MW874147	*Termitomyces* sp. T8	SRX10474843
Termitomyces T99 pT99 1	MW874167	*Termitomyces* sp. T99	SRX10767243
Termitomyces T99 pT99 2	MW874145	*Termitomyces* sp. T99	SRX10767243
Termitomyces titanicus pTT	MW874170	*Termitomyces titanicus*	SRX10337357
Termitomyces titanicus pTT mitochondrion	MW874156	*Termitomyces titanicus*	SRX10337357
Trametes versicolor pTV	MW874121	*Trametes versicolor*	n.a.

Assembly and plasmid detection

We assembled paired-end Illumina sequencing reads using SPAdes version 3.5.0 with default settings (Bankevich et al. 2012). We considered contigs potentially representing linear plasmids using their characteristic invertron-like structure. Contigs were sorted according to length and all contigs between five and fifteen kb in size (all known invertron plasmids fall into this range) had their open reading frames predicted using the mold mitochondrial code. Any contigs with two large (1kb+) open reading frames facing each other were considered potential plasmids. The two open reading frames were then checked for homology to known plasmids using BLASTp to verify they were plasmid-associated RNA and DNA polymerases. If so, the contig was considered to represent an invertron plasmid. After compiling all plasmids, we used BLASTn to screen all assemblies again to search for plasmid sequences that perhaps had not been assembled contiguously and thereby had escaped detection, for example where the polymerases were split across two contigs.

Most plasmids gathered in this way most likely represent autonomous, non-inserted plasmids as they were capped on each end by an inverted repeat and their contig did not extend into a flanking region of mtDNA. Generally, inserted plasmids degenerate over evolutionary time and the polymerase coding sequences break down through stop codon accumulation. However, it is possible for a relatively recent plasmid insertion to be mistaken for an autonomous plasmid. In three *Termitomyces* samples (*T. titanicus, T. bulborhizus,* and *T. sp.* K1Ac) we found near-intact mitochondrial plasmid inserts. These could be distinguished from autonomous plasmids based on extended flanking regions beyond the terminal inverted repeats that shared homology with mtDNA. This shows that even recent insertions of complete plasmid sequences may be distinguished from autonomous copies by the flanking regions. We included these inserted plasmids in the phylogeny as the DNA and RNA polymerases were still intact.

Sequence alignment and phylogenetic reconstruction

We aligned the translated amino acid sequences of the RNA and DNA polymerases separately using MAFFT (Katoh et al. 2002) with default parameters. The nucleotide sequences were translated to amino acids using the mold mitochondrial genetic code before alignment. We used the Gblocks online server (http://molevol.cmima.csic.es/castresana/Gblocks_server.html, Gblocks version 0.91b) (Castresana 2000) to remove poorly aligned regions, using the following settings: 'allow smaller final blocks', 'allow gap positions within final blocks', and 'allow less strict flanking positions'. We then concatenated the trimmed alignments of both polymerases.

We used IQ-TREE for phylogenetic reconstruction (Nguyen et al. 2015). ModelFinder (Kalyaanamoorthy et al. 2017) found the best-fitting model for both the DNA and RNA

polymerase alignment to be LG+F+I+G4. We ran IQ-TREE including 10000 ultrafast bootstraps (Hoang et al. 2018). Nodes with bootstrap support values below 95 were collapsed into polytomies. We included a midpoint root as an approximate rooting since there is no known closely related sister group to these linear plasmids. Some plant plasmids are suspected to form a sister group but that relationship was not significantly supported in our trial phylogeny (data not shown). Previous studies have used phage polymerases as outgroup (Andrade & Góes-Neto 2015) but we considered these too divergent to be informative.

Detection of DNA exchange between mtDNA and plasmid

We performed reciprocal BLAST searches of curated, complete mitochondrial genomes of twelve Lyophyllaceae species (Nieuwenhuis et al. 2019) to all plasmid sequences (settings: Megablast, match mismatch 1 2, gap cost linear, word size 28). We considered hits significant if they exceeded 50bp in length and had an E-value <0.001.

To verify tRNA sequences and examine the effect of mutations on their secondary structure, we used the tRNA-Scan web server (Chan & Lowe 2019). We set sequence source to 'Other Mitochondrial' and genetic code to 'Mold & Protozoan Mito', with all other settings as default.

Results

We collected 55 plasmid sequences from 54 whole-genome sequence assemblies. With an additional 18 plasmid sequences found on GenBank, our final data set comprised 73 plasmid sequences from 46 host strains. The highest number of different plasmid strains found in a single host was five for *Termitomyces* sp. T123. We observed intraspecific variation of plasmid presence among host species for which multiple strains were analysed: we detected up to three different types of plasmid in *Termitomyces cryptogamus* ranging from one in strain J132, to three in strain T8, with two in the five other strains of this species.

Phylogeny

As expected based on previous research (Andrade & Góes-Neto 2015; Griffiths 1995; Van Der Gaag et al. 1998), the plasmid phylogeny is incongruent with the host fungal phylogeny (Figure 1), with many clades consisting of plasmids from ascomycete and basidiomycete hosts for example. Deep nodes have poor support and are mostly collapsed to polytomies, due to presumed long evolutionary distance between plasmid clades and limited alignment length resulting in overall poor homology. However, we obtained strong support for several clades of relatively closely related plasmids. Many of

these clades consist of a mix of Lyophyllaceae-associated plasmids and plasmids of more distantly related hosts.

Plasmid-mtDNA integrations

Comparison using BLAST of mitochondrial DNA to all plasmid sequences revealed numerous homologous sequences between plasmids and mtDNA (Figure 1). In most cases, these potential inserts were small and highly degraded, with the polymerase genes interspersed with numerous stop codons. In host mtDNA infected with autonomous plasmids, inserts often do not share homology with those autonomous plasmids, suggesting the inserts derive from plasmids that have since disappeared from that host. Inserts were also often found in species for which we found no autonomous plasmids, again indicating past presence of plasmids.

Transfer of mtDNA to plasmid

We detected several instances of mitochondrial DNA of the fungal host integrated in plasmid DNA (Figure 1). In plasmids pT132_1/pJ132/pT99_2/pT8_3/pT28_3/pt60a_1/pT61_1 (all from the species *Termitomyces cryptogamus*), we detected a sequence homologous to tRNA-Arg(TCG) from mitochondrial genomes of *Termitomyces* and close relatives (Figure 2). In particular, flanking regions of this plasmid sequence matched flanking regions of tRNA-Arg(TCG) in the mtDNA of the sister group of *T. cryptogamus*, *T.* sp. T123. Furthermore, in the mitochondrion of *T.* sp. T132, this mtDNA region lacks the tRNA-Arg(TCG) gene as well as homologous sequences of the flanking regions. This suggests that in the ancestor of *T. cryptogamus*, this part of mtDNA was copied to a plasmid, after which the mitochondrial copy was lost (Figure 3). Although we found many mismatches between the plasmid pT132_1 and the mtDNA of *T.* sp. T123 within the flanking regions, we found no mutations in the tRNA itself. As we do not know the exact timing of the tRNA transfer to the plasmid following the divergence of *T.* sp. T132 and *T.* sp. T123, it is unclear if these mutations mostly occurred in the host mtDNA prior to transfer, or in the plasmid after transfer.

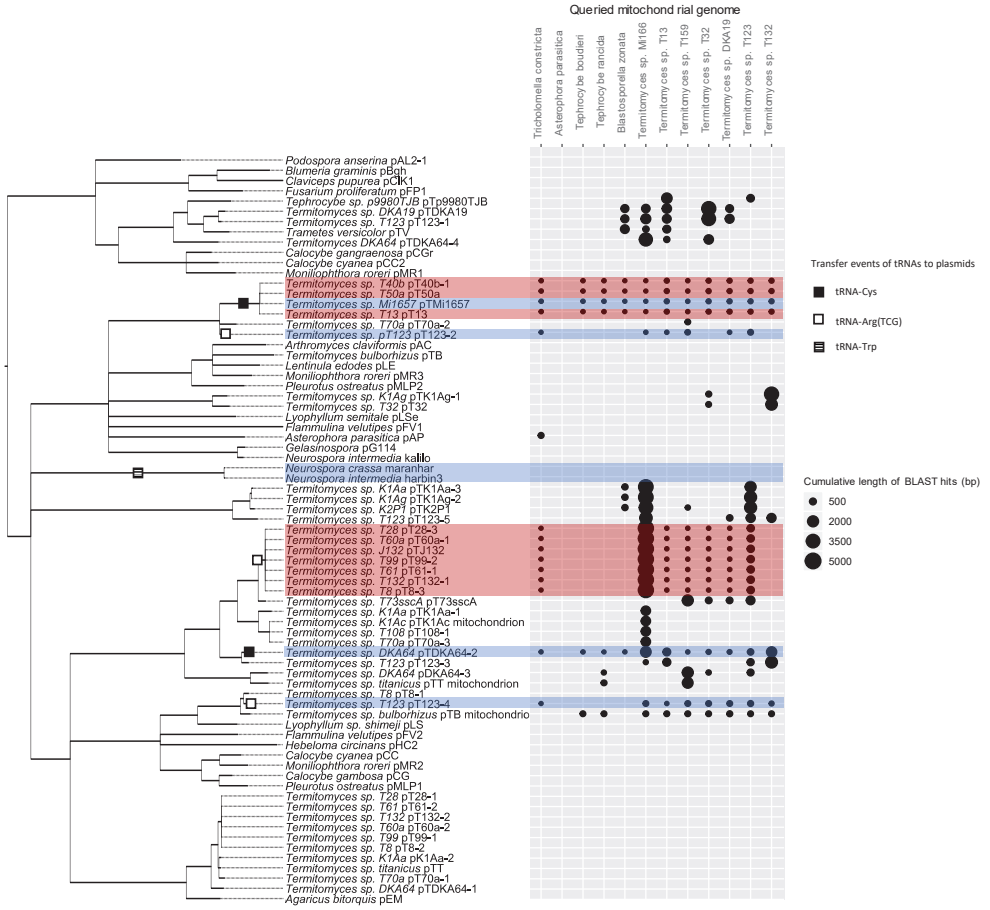


Figure 1: Midpoint-rooted maximum likelihood phylogeny of fungal linear plasmids. Nodes with less than 95 ultrafast bootstrap support values were collapsed. Plasmids are named after the host species or strain in which they are found. In strains with multiple plasmids the name is extended with a number to differentiate them. Plasmids that were found as intact mitochondrial insertions are indicated with the word 'mitochondrion'. The graph on the right indicates the presence and cumulative length of homologous regions found between each plasmid and twelve mitochondrial genomes of the Lyophyllaceae. Plasmids with adopted mitochondrial tRNA genes are color-coded in red (in case the same tRNA gene was lost in the host mtDNA) and blue (in case the same tRNA gene is still present in the host mtDNA). Inferred independent transfer events of tRNA genes to plasmids are indicated by squares on the phylogeny (black for tRNA-Arg(TCG), white for tRNA-Cys, and striped for tRNA-Trp).

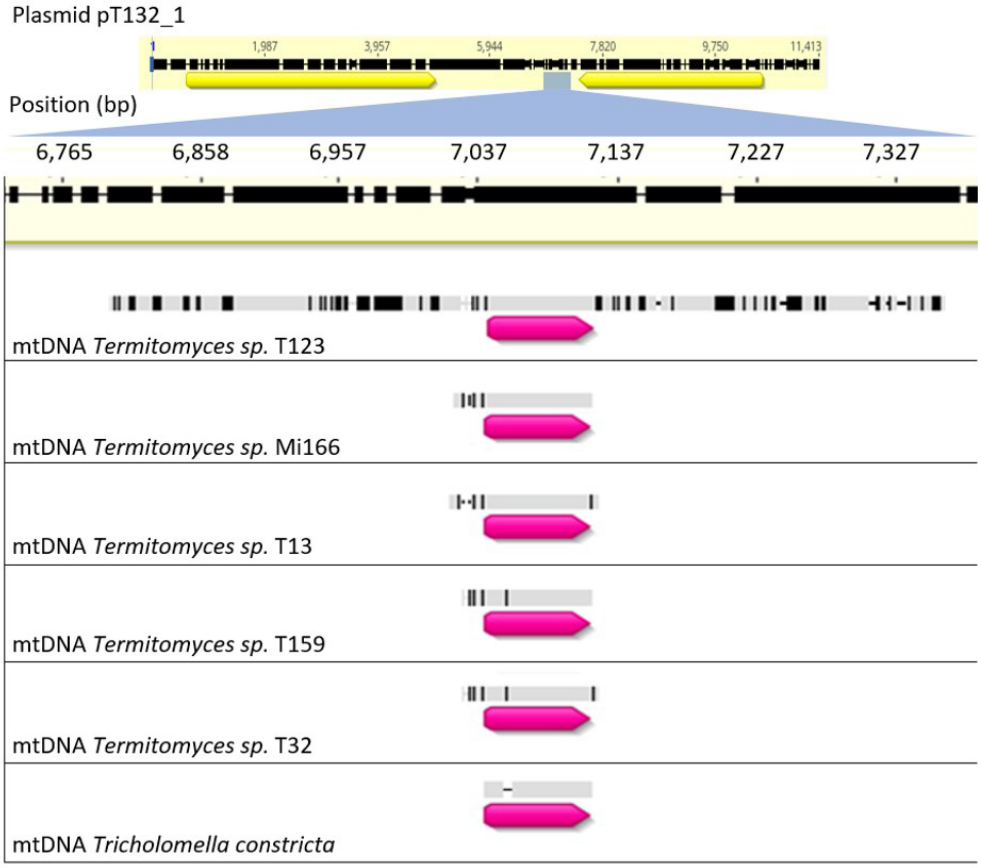


Figure 2: Homology of tRNA-Arg(TCG) of several Lyophyllaceae mitochondrial genomes and a region of plasmid pT132_1 detected by Megablast. A roughly 600 base pair section of pT132_1 located between the DNA and RNA polymerase is shown (black bar, top). Below are shown homologous regions from the mtDNA of several Lyophyllaceae species (grey bars, black sections indicate sequence divergence). The location and relative size of tRNA-Arg(TCG) for each species is indicated by the pink arrow. While most species only show a significant match for the exact location of tRNA-Arg(TCG), *Termitomyces* sp. T123 shares significant homology with the plasmid for the flanking regions as well (up to around 300 base pairs upstream and downstream). Note that the host species carrying the plasmid, *Termitomyces* sp. T132, showed no significant homology between the plasmid and its mtDNA.

Termitomyces sp. T123 is associated with five different linear plasmids, all of which have been detected in the same isolate. Two of these plasmids contain a copy of the tRNA-Arg(TCG) gene but have acquired it independently from each other and from the plasmid

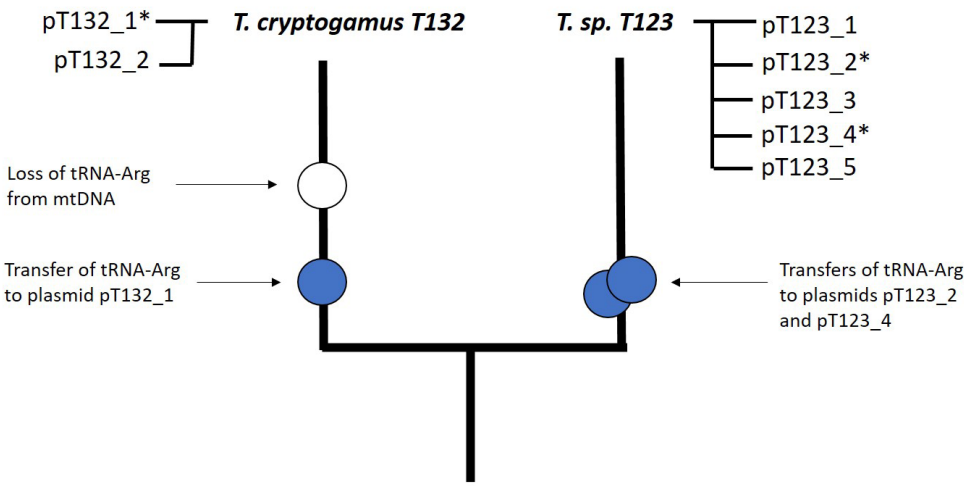


Figure 3: Multiple transfers of mitochondrial tRNA genes to linear plasmids in two closely related species of Termitomyces. *Termitomyces cryptogamus* (here represented by strain T132) and T. sp. T123 are sister species, and both are associated with multiple linear plasmids. At least three independent transfer events of a mitochondrial tRNA-Arg gene occurred between these two fungal lineages, with one subsequent loss of the mitochondrial tRNa-Arg copy in the ancestor of *T. cryptogamus.* A hypothetical sequence of these events is shown on the phylogeny. Plasmids encoding a tRNA-Arg copy are indicated with an asterisk.

found in T. *cryptogamus* (Figure 1). The same tRNA-Arg(TCG) gene is also still found in the mtDNA of *T.* sp. T123, so the host mtDNA is not dependent on the plasmid copies for this anticodon. In plasmid pT123_2, the tRNA-Arg(TCG) carries mutations that alter its secondary structure according to a structural prediction made using tRNA-Scan. In plasmid pT123_4, no mutations have occurred in the tRNA-Arg(TCG) gene relative to its host mitochondrial counterpart, but whether this is due to selection pressure or due to this copy being a recent acquisition is unclear. pT123_4 also contains a sequence matching the first 48 amino acids of the nad2 gene, with several mutations at the 5'-end.

Plasmids pT50a/pT40b/pTMi1657/pT13 possess a copy of tRNA-Cys. This gene is missing in the mtDNA of *Termitomyces* sp. T13/sp. T50a/sp. T40b, indicating another transfer of a tRNA followed by mitochondrial loss in some strains carrying this plasmid but not in all. In *Termitomyces sp.* DKA64, we also found a plasmid with a copy of tRNA-Cys, with no loss of this tRNA in the mtDNA of this strain. As this plasmid is not closely related to the other four, this likely represents an independent transfer event.

We also found that the maranhar plasmid associated with *Neurospora crassa* contained a tRNA-Trp gene. A copy of this same tRNA-Trp gene has been detected in the *Neurospora*-associated circular plasmids, *Varkud* and *Mauriceville* (Akins *et al.* 1989).

Codon frequency of captured tRNA***s***

We observe at least six independent adoption events of tRNA genes by plasmids, once in *Neurospora* and five times in *Termitomyces*. However, of the 24 different tRNA genes present in almost all fungal mtDNA, only three feature in these adoption events: three cases of tRNA-Arg(TCG), two of tRNA-Cys and one tRNA-Trp. We hypothesized that the probability of adoption of a tRNA gene by a plasmid may depend on the frequency of the corresponding codon found in the mitochondrial genes. We therefore calculated the average codon use for the twelve Lyophyllaceae mitochondrial genomes (Figure 4). We found that the tRNA genes transferred to plasmids in Lyophyllaceae (tRNA-Arg(TCG) and tRNA-Cys) rank the lowest in terms of their corresponding codon use in mitochondrial genes.

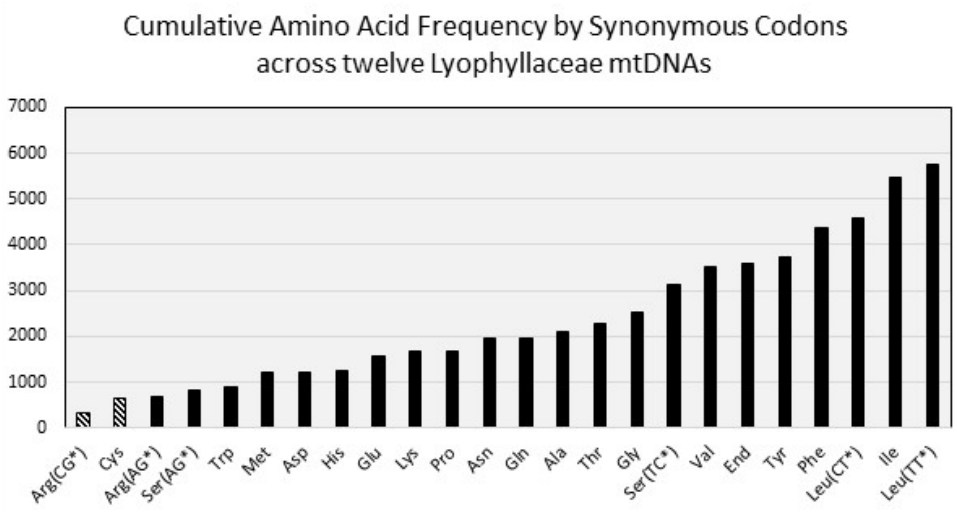


Figure 4: Cumulative amino acid frequency of the fourteen principal mitochondrial genes (atp6, atp8, atp9, cob, cox1, cox2, cox3, nad1, nad2, nad3, nad4, nad4L, nad5, nad6) for twelve Lyophyllaceae mtDNAs. We grouped amino acids by synonymous codons, which correspond to the anticodons carried by mitochondrial tRNAs. The amino acid codons for which we observe transfer of the corresponding tRNA gene to a mitochondrial plasmid (in the Lyophyllaceae) are shown with diagonal stripes.

Discussion

Plasmid diversity and evolution

Invertron plasmids are thought to be the most abundant group of mitochondrial plasmids in fungi (Griffiths 1995). Of the 54 Lyophyllaceae strains sampled for this study, over half (28) had at least one associated plasmid strain. Studies in *Podospora* (Van Der Gaag et al. 1998) and *Neurospora* (Arganoza et al. 1994; Maas et al. 2005) have previously shown intraspecific variation in plasmid presence of fungal isolates. Consistent with this, we found intraspecific variation in plasmid infection for the seven strains we analysed of one *Termitomyces* species, *T. cryptogamus* (*T.* sp. J132, sp. T132, sp. T8, etc.). As such it is likely that in at least some of the species for which our sequenced strains showed no plasmids in this study, strains in the wild exist that do have plasmids.

Our phylogenetic tree (Figure 1) combines the 54 newly discovered plasmids from the Lyophyllaceae with 18 that were previously reported in other fungi, and shows significant support for clades incongruent with the host-fungal phylogeny. Several clades contain a mixture of plasmids derived from ascomycetes and basidiomycetes. The majority of plasmids analysed in this study were found in the genus *Termitomyces*, and many of these cluster into different clades with plasmids from other fungi. These incongruencies with the host phylogeny can partly be explained by a long history of differential loss of plasmids among hosts. However, considering the fast mutation rate of these plasmids (Warren et al. 2016) and the relatively low genetic divergence between plasmids of different host genera, horizontal transmission of plasmids between different fungal species probably plays a significant part as well.

A couple of plasmids we discovered in two *Calocybe* species (pCGr and pCC2) did not conform to the expected invertron structure, with the DNA polymerase and RNA polymerase genes positioned on the same strand. Both plasmids were still capped on each end by terminal inverted repeats. In our phylogenetic reconstruction (Figure 1), these plasmids form a clade with one of the plasmids associated with *Moniliophthora roreri* (pMR1), which also has a non-invertron structure (Costa et al. 2012).

Plasmid insertions in mtDNA

Sequence comparison of twelve Lyophyllaceae mitochondrial genomes to our plasmid data set revealed numerous and sizeable plasmid insertions in mtDNA, even in species for which no autonomous plasmid was detected (Figure 1). Furthermore, in species with known plasmid infections, mitochondrial inserts were often not homologous to the autonomous plasmid. Since most plasmid inserts were fragmented and degraded, these findings suggest most inserts represent old insertion events that occurred between the

ancestral mtDNA and an autonomous plasmid that in many cases no longer occurs in the isolate.

Transfer of mtDNA to plasmids

Transfer of mitochondrial DNA to linear and circular plasmids has previously been described in fungi (Akins *et al.* 1986, Akins *et al.* 1989, Kempken 1989, Mohr *et al.* 2000) and in plants (Leon *et al.* 1989). In circular plasmids, such transfers may occur by the same reverse transcription process that drives plasmid replication (Chiang & Lambowitz 1997). The transfers we identified in invertron plasmids may also result from replication errors by the plasmid, or a more general process of insertion. The replication process of invertron plasmids is thought to involve a protein primer, using the proteins covalently bound to the terminal inverted repeats. This differs significantly from circular plasmids, which are thought to be able to replicate without a primer but initiate close to a tRNA-like structure.

We found at least two separate occasions of mitochondrial loss of a tRNA gene following the transfer of a copy of that gene to a plasmid (Figure 1). Leon *et al.* found that in the maize relative *Teosinte* the transfer of a tRNA-Trp gene coincided with its loss in the mtDNA. Warren *et al.* (2016) suggested that transfer of mtDNA to plasmids contributes to accelerated sequence evolution of such genes due to the higher mutation rate of plasmids. Our findings of conserved tRNA sequences in plasmids show that such sequences may in fact remain relatively free from mutations when the mitochondrial copy is lost and the transferred sequence encodes an essential function.

Mitochondria import many tRNAs from the nucleus, which could be an alternative explanation for mitochondrial gene loss from mtDNA. However, the sudden loss of tRNAs co-occurring with the appearance of a functional copy in a mitochondrial plasmid that appears to be fixed in the population suggests the tRNA function has transferred to the plasmid and not to the nucleus. In addition, we only observe loss of mitochondrial tRNA genes in Lyophyllaceae mtDNA when a copy of that tRNA gene is present on an autonomous plasmid. If tRNA transcripts were imported from the nucleus, loss of a mitochondrial tRNA would occasionally be expected in the absence of a plasmid copy.

Ferandon et al. (Ferandon et al. 2008) reported a tRNA-Met gene located on a linear plasmid that had integrated in the mtDNA of its host, the fungus *Agrocybe agerita*, and suggested the tRNA gene may have been captured by the plasmid prior to integration. Mitochondrial insertion of plasmids carrying mitochondrial tRNA genes or other mitochondrial genes poses an interesting evolutionary process. First of all, plasmids can be horizontally transferred between different species, creating the potential for indirect horizontal transfer of mtDNA between these species. Second, following mitochondrial loss of a tRNA gene that transferred to a plasmid, the plasmid could reintroduce the tRNA

gene to the mtDNA through insertion. This would likely alter the location of the tRNA gene in the mtDNA. It would also nullify any selective benefit the plasmid enjoyed from the transferred tRNA gene if the mtDNA is able to recover the tRNA function.

Plasmid pT123_4 also contained portions of nad2, a mitochondrial protein-coding gene. This shows plasmids are capable of capturing other parts of mtDNA besides tRNA genes, but since captured sequences are generally limited in size, capturing a complete functional gene may be restricted to tRNA genes and small genes like atp8, atp9 and nad4L.

All tRNAs found in plasmids code for low-frequency codons

We observe at least five independent acquisitions of tRNA genes by plasmids in the genus *Termitomyces*, twice of tRNA-Cys and three times of tRNA-Arg(TCG). Why do we only observe these tRNA genes in plasmids and no others? It is possible that the capture of mtDNA by plasmids is dependent on certain sequence motifs found only nearby these tRNA genes. However, the intergenic sequences are not well conserved between *Termitomyces* species making this unlikely. Looking at the codon usage of mitochondrial genes (Figure 4), the codons that tRNA-Cys and tRNA-Arg(TCG) encode anticodons for are the least frequently used of all codons. If the production of tRNA transcripts by plasmids is lower than that of mitochondrial tRNA genes (as they are not adapted for their transcription), it is possible that plasmids can only meet the mitochondrial demand of rare anticodons. Such captured tRNA genes can then be selectively beneficial if the mitochondrial copy is lost or mitochondrial transcription of that tRNA gene is otherwise inhibited. However, this would not explain why we also find only low-frequency codon-coding tRNA genes in plasmids when the host mtDNA has not lost its own copy. Fungal mtDNA is thought to be generally transcribed in large polycistronic transcripts (Kolondra et al. 2015), which might result in roughly equal transcription rates for tRNA genes. It is possible that low-frequency codons have more unused transcripts available for integration by a plasmid, increasing their chance of transfer to a plasmid.

Capture of tRNA genes coding for codons commonly used in the host mtDNA was shown by Akins et al. (Robert A Akins et al. 1989) in *Neurospora crassa,* where circular plasmids were found containing tRNA-Trp, tRNA-Val and tRNA-Gly. These three tRNA genes code for codons that appear in the mtDNA at relatively high frequencies (Figure 4). This shows it is possible for plasmids to adopt high-demand tRNA genes, but for some reason we did not observe any such cases in linear plasmids.

Conclusions

We recovered 55 mitochondrial linear plasmid sequences from the Lyophyllaceae and analysed DNA transfer between these plasmids and their host mtDNA. We observed several independent transfers of complete mitochondrial tRNA genes to linear plasmids. In two cases, this transfer coincided with a loss of the tRNA gene in the host mtDNA, potentially resulting in genetic addiction of the host to the plasmid. We also show that tRNA genes transferred to plasmids tend to code for rare mitochondrial codons. Our results show that the interaction between fungal mtDNA and linear mitochondrial plasmids is much more dynamic than previously thought. The potential addiction of mtDNA to a plasmid through the exchange of a tRNA gene shows how a mutualistic association can arise abruptly between a selfish genetic element and its host genome.

References

Akins R A, Kelley RL, Lambowitz AM. 1989. Characterization of mutant mitochondrial plasmids of Neurospora spp. that have incorporated tRNAs by reverse transcription. Mol. Cell. Biol. 9:678–691. doi: 10.1128/mcb.9.2.678.

Akins Robert A, Kelley RL, Lambowitz AM. 1989. Characterization of Mutant Mitochondrial Plasmids of Neurospora spp . That Have Incorporated tRNAs by Reverse Transcription. 9:678–691.

Andrade BS, Góes-Neto A. 2015. Phylogenetic analysis of DNA and RNA polymerases from a Moniliophthora perniciosa mitochondrial plasmid reveals probable lateral gene transfer. Genet. Mol. Res. 14:14105–14114. doi: 10.4238/2015.October.29.30.

Arganoza MT, Min J, Hu Z, Akins RA. 1994. Distribution of seven homology groups of mitochondrial plasmids in Neurospora: evidence for widespread mobility between species in nature. Curr. Genet. 26:62–73. doi: 10.1007/BF00326306.

Bai X, Debaud JC, Schründer J, Meinhardt F. 1997. The ectomycorrhizal basidiomycete Hebeloma circinans harbors a linear plasmid encoding a DNA- and RNA polymerase. J. Gen. Appl. Microbiol. 43:273–279. doi: 10.2323/jgam.43.273.

Bankevich A et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455–477. doi: 10.1089/cmb.2012.0021.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17:540–552. doi: 10.1093/oxfordjournals.molbev.a026334.

Chan PP, Lowe TM. 2019. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. Methods Mol. Biol. 1962:1–14. doi: 10.1007/978-1-4939-9173-0_1.

Chiang CC, Lambowitz AM. 1997. The Mauriceville retroplasmid reverse transcriptase initiates cDNA synthesis de novo at the 3' end of tRNAs. Mol. Cell. Biol. 17:4526 LP – 4535. doi: 10.1128/MCB.17.8.4526.

Costa GGL et al. 2012. The mitochondrial genome of Moniliophthora roreri, the frosty pod rot pathogen of cacao. Fungal Biol. 116:551–562. doi: 10.1016/j.funbio.2012.01.008.

Court DA, Bertrand H. 1992. Genetic organization and structural features of maranhar, a senescence-inducing linear mitochondrial plasmid of Neurospora crassa. Curr. Genet. 22:385–397. doi: 10.1007/BF00352440.

Ferandon C, Chatel SEK, Castandet B, Castroviejo M, Barroso G. 2008. The Agrocybe aegerita mitochondrial genome contains two inverted repeats of the nad4 gene arisen by duplication on both sides of a linear plasmid integration site. Fungal Genet. Biol. 45:292–301. doi: 10.1016/j.fgb.2007.10.012.

Galligan JT, Kennell JC. 2007. Retroplasmids: Linear and Circular Plasmids that Replicate via Reverse Transcription. In: Microbial Linear Plasmids. Meinhardt, F & Klassen, R, editors. Springer Berlin Heidelberg: Berlin, Heidelberg pp. 163–185. doi: 10.1007/7171_2007_098.

Giese H, Lyngkjaer MF, Stummann BM, Grell MN, Christiansen SK. 2003. Analysis of the structure and inheritance of a linear plasmid from the obligate biotrophic fungus Blumeria graminis f. sp. hordei. Mol. Genet. Genomics. 269:699–705. doi: 10.1007/s00438-003-0876-5.

Griffiths AJF et al. 1990. Heterokaryotic transmission of senescence plasmid DNA in Neurospora. Curr. Genet. 17:139–145. doi: 10.1007/BF00312859.

Griffiths AJF. 1995. Natural plasmids of filamentous fungi. Microbiol. Rev. 59:673–685. doi: 10.1128/mmbr.59.4.673-685.1995.

Hermanns J, Osiewacz HD. 1992. The linear mitochondrial plasmid pAL2-1 of a long-lived Podospora anserina mutant is an invertron encoding a DNA and RNA polymerase. Curr. Genet. 22:491–500. doi: 10.1007/BF00326415.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. 35:518–522. doi: 10.1093/molbev/msx281.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods. 14:587–589. doi: 10.1038/nmeth.4285.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066. doi: 10.1093/nar/gkf436.

Kempken F, Meinhardt F, Esser K. 1989. In organello replication and viral affinity of linear, extrachromosomal DNA of the ascomycete Ascobolus immersus. Mol. Gen. Genet. 218:523–530. doi: 10.1007/BF00332419.

Kempken F. 1989. A tRNA-like structure from a linear plasmid of the fungus Ascobolus immersus. Nucleic Acids Res. 17:6740. doi: 10.1093/nar/17.16.6740.

Kempken F. 1995. Horizontal transfer of a mitochondrial plasmid. MGG Mol. Gen. Genet. 248:89–94. doi: 10.1007/BF02456617.

Kennell JC, Wang H, Lambowitz AM. 1994. The Mauriceville plasmid of Neurospora spp. uses novel mechanisms for initiating reverse transcription in vivo. Mol. Cell. Biol. 14:3094–3107. doi: 10.1128/mcb.14.5.3094.

Kim EK, Jeong JH, Youn HS, Koo YB, Roe JH. 2000. The terminal protein of a linear mitochondrial plasmid is encoded in the N-terminus of the DNA polymerase gene in white-rot fungus Pleurotus ostreatus. Curr. Genet. 38:283–290. doi: 10.1007/s002940000157.

Kobayashi I. 2004. Genetic Addiction: a Principle of Gene Symbiosis in a Genome. Plasmid Biol. 105–144. doi: https://doi.org/10.1128/9781555817732.ch6.

Kolondra A, Labedzka-Dmoch K, Wenda JM, Drzewicka K, Golik P. 2015. The transcriptome of Candida albicans mitochondria and the evolution of organellar transcription units in yeasts. BMC Genomics. 16:827. doi: 10.1186/s12864-015-2078-z.

Láday M, Stubnya V, Hamari Z, Hornok L. 2008. Characterization of a new mitochondrial plasmid from Fusarium proliferatum. Plasmid. 59:127–133. doi: 10.1016/j.plasmid.2007.11.006.

Leon P, Walbot V, Bedinger P. 1989. Molecular analysis of the linear 2.3 kb plasmaid of maize mitochondria: apparent capture of tRNA genes. Nucleic Acids Res. 17:4089–4099.

Maas MFPM, Boer HJ d., Debets AJM, Hoekstra RF. 2004. The mitochondrial plasmid pAL2-1 reduces calorie restriction mediated life span extension in the filamentous fungus Podospora anserina. Fungal Genet. Biol. 41:865–871. doi: https://doi.org/10.1016/j.fgb.2004.04.007.

Maas MFPM, van Mourik A, Hoekstra RF, Debets AJM. 2005. Polymorphism for pKALILO based senescence in Hawaiian populations of Neurospora intermedia and Neurospora tetrasperma. Fungal Genet. Biol. 42:224–232. doi: https://doi.org/10.1016/j.fgb.2004.11.004.

Mohr S, Wanner LA, Bertrand H, Lambowitz AM. 2000. Characterization of an unusual tRNA-like sequence found inserted in a Neurospora retroplasmid. Nucleic Acids Res. 28:1514–1524. doi: 10.1093/nar/28.7.1514.

Nakai R, Sen K, Kurosawa S, Shibai H. 2000. Basidiomycetous fungus Flammulina velutipes harbors two linear mitochondrial plasmids encoding DNA and RNA polymerases. FEMS Microbiol. Lett. 190:99–102. doi: 10.1111/j.1574-6968.2000.tb09269.x.

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. 32:268–274. doi: 10.1093/molbev/msu300.

Nieuwenhuis M et al. 2019. Enrichment of G4DNA and a Large Inverted Repeat Coincide in the Mitochondrial Genomes of Termitomyces. Genome Biol. Evol. 11. doi: 10.1093/gbe/evz122.

Oeser B, Tudzynski P. 1989. The linear mitochondrial plasmid pClK1 of the phytopathogenic fungus Claviceps purpurea may code for a DNA polymerase and an RNA polymerase. Mol. Gen. Genet. MGG. 217:132–140. doi: 10.1007/BF00330952.

Robison MM, Royer JC, Horgen PA. 1991. Homology between mitochondrial DNA of Agaricus bisporus and an internal portion of a linear mitochondrial plasmid of Agaricus bitorquis. Curr. Genet. 19:495–502. doi: 10.1007/BF00312742.

Shiu-Shing Chan B, 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. doi: 10.1007/BF00326237.

Tudzynski P, Stahl U, Esser K. 1980. Transformation to senescence with plasmid like DNA in the ascomycete Podospora anserina. Curr. Genet. 2:181–184. doi: 10.1007/BF00435683.

Van Der Gaag M, Debets AJM, Osiewacz HD, Hoekstra RF. 1998. The dynamics of pAL2-1 homologous linear plasmids in Podospora anserina. Mol. Gen. Genet. 258:521–529. doi: 10.1007/s004380050763.

van Diepeningen AD, Debets AJM, Slakhorst SM, Hoekstra RF. 2008. Mitochondrial pAL2-1 plasmid homologs are senescence factors in Podospora anserina independent of intrinsic senescence. Biotechnol. J. 3:791–802. doi: 10.1002/biot.200800005.

Warren JM, Simmons MP, Wu Z, Sloan DB. 2016. Linear Plasmids and the Rate of Sequence Evolution in Plant Mitochondrial Genomes. Genome Biol. Evol. 8:364–374. doi: 10.1093/gbe/evw003.

Xu Y, Yang S, Turitsa I, Griffiths A. 1999. Divergence of a Linear and a Circular Plasmid in Disjunct Natural Isolates of the Fungus Neurospora. Plasmid. 42:115–125. doi: https://doi.org/10.1006/plas.1999.1420.

Yang R et al. 2017. The complete mitochondrial genome of the widely cultivated edible fungus Lentinula edodes. Mitochondrial DNA Part B. 2:13–14. doi: 10.1080/23802359.2016.1275839.

Yuewang W, Yang X, Griffiths AJF. 1996. Structure of a Gelasinospora linear plasmid closely related to the kalilo plasmid of Neurospora intermedia. Curr. Genet. 29:150–158. doi: 10.1007/s002940050030.

CHAPTER 5

5

Mitochondrial inheritance in the basidiomycete *Termitomyces cryptogamus*

Mathijs Nieuwenhuis, Duur K. Aanen

Abstract

Abstract

Mitochondrial inheritance is generally thought to be uniparental, possibly to avoid heteroplasmy and recombination of mitochondrial DNA (mtDNA). Species that appear to have biparental transmission of mtDNA are of interest as such exceptions to the rule may reveal the effects of cytoplasmic mixing on the stability of the nuclear-mitochondrial mutualism. In basidiomycetes, uniparental transmission is generally achieved by reciprocal nuclear exchange (or nuclear migration) between compatible homokaryons during sex. Nonetheless biparental inheritance and mitochondrial recombinants are observed in several basidiomycetes such as *Agaricus bisporus,* which lacks nuclear migration. Species of the basidiomycete genus *Termitomyces* show similarities in mating to *Agaricus bisporus,* with an apparent lack of nuclear migration. In this study, we compare mtDNA sequences from the interaction zones of mated *Termitomyces cryptogamus* strains and heterokaryotic strains from wild populations to see if recombination of mtDNA is also observed in this genus. Our results indicate recombination of mtDNA in *Termitomyces cryptogamus* is restricted to self-splicing introns that can recombine separately from other mtDNA sequences through homing-endonuclease activity. Our results show the importance of taking homing endonucleases into account when studying recombination of mtDNA in fungi.

Introduction

Mitochondrial DNA (mtDNA) is generally inherited uniparentally and recombination of mtDNA is relatively uncommon. This uniparental non-recombinant mode of inheritance, often combined with bottlenecking of mtDNA which further reduces genetic variation, facilitates a cytoplasmic state dominated by a single mitochondrial haplotype (homoplasmy). As mitochondria are endosymbionts, homoplasmy prevents competition between divergent mitochondrial genomes that could be detrimental to the host cell (Christie et al. 2015). Nonetheless biparental inheritance and recombination of mtDNA have been reported in various organisms including fungi (De La Bastide & Horgen 2003; Barr et al. 2005), as well as (transient) heteroplasmy (Wang et al. 2017; Saville et al. 1998). This raises the question why cytoplasmic inheritance appears to be strictly regulated to avoid heteroplasmy and recombination in some organisms, while in others it appears not to be so.

Fungi showcase a wide variety of sexual reproductive strategies including use of multiple mating loci, mating type switching, selfing and outcrossing (known as homo- and heterothallism, respectively). In addition, some fungi reproduce sexually by fusion of multicellular hyphae while others fuse through yeast-like single cells. This diversity in mating strategies results in a parallel diversity of cytoplasmic inheritance, with uniparental and biparental inheritance found in different fungi, including in some cases mtDNA recombination (Barr et al. 2005; Wang et al. 2017; van Diepeningen et al. 2010; Xu et al. 2009; Fritsch et al. 2014). However, many cases of mtDNA recombination reported in fungi are restricted to recombination of sequences surrounding endonuclease genes, which are often found in type 1 introns (van Diepeningen et al. 2010; Xu et al. 2009) but can also occur as standalone endonuclease genes (Wu & Hao 2019). Endonuclease genes and their adjacent mtDNA sequences recombine separately from the rest of the mtDNA through RNA-mediated self-splicing and gene conversion (Haugen et al. 2005; Delahodde et al. 1989; Mueller et al. 1996). Therefore, recombination of mtDNA limited to endonucleases and their flanking regions is not indicative of 'normal' recombination between different mtDNAs, but it could indicate cytoplasmic mixing (Gogarten & Hilario 2006).

In most filamentous basidiomycetes, mating involves the transfer of nuclei through a short-lived fusion zone between mating partners. Mitochondria do not migrate along with nuclei across the fusion zone (May & Taylor 1988; Anderson & Kohn 2014). Each fertilized hypha therefore maintains its cytoplasm after mating, effectively resulting in doubly uniparental inheritance, and a mosaic of heterokaryons that contain similar nuclei but different parental mitochondria. Within the fusion zone, there is cytoplasmic mixing and potential recombination of mtDNA (Hickey et al. 2002), but these cells are generally terminated shortly after hyphal anastomosis (Ainsworth & Rayner 1986; Rayner 1991).

The basidiomycete *Termitomyces* is thought to have limited or no nuclear migration during mating, since the heterokaryon forms in the central mass of the fusion zone rather than on either side of it as is common in fungi with nuclear migration (Nobre et al. 2014). If nuclear migration is absent and the heterokaryon grows directly from the fusion zone, *Termitomyces* may show biparental inheritance and frequent recombination of mtDNA. These assumptions are supported by evidence from the basidiomycete genus *Agaricus,* in which nuclear migration is also either limited (Hintz et al. 1988) or absent (Elliott 1978; Jin et al. 1992). Consistent with the expected increased scope for mtDNA recombination due to more frequent heteroplasmy, some studies have reported mitochondrial recombination in some crossings of *Agaricus* strains or select wild populations (Xu et al. 2013; De La Bastide & Horgen 2003).

In this study, we investigated whether mitochondrial recombination occurs in *Termitomyces* following mating under laboratory conditions, as well as in natural populations. For the experimental matings, we isolated single-spore cultures from conidia isolated from the fusion zones and used Sanger sequencing of three mitochondrial loci (one inside a type 1 intron) to test for recombination. For the natural populations, we compared whole-genome sequencing data for six strains obtained from the wild to test 47 polymorphisms for signs of recombination.

Methods

Strain and marker selection

We used two sexually compatible homokaryon strains (T132 and P5) of a *Termitomyces* species, *T. cryptogamus* (Van de Peppel et al., unpublished) associated with the termite species *Macrotermes natalensis*. A previous study has reconstructed the annotated mitochondrial-genome sequence for T132 (GenBank accession: MG783568.1) (Nieuwenhuis et al. 2019). We mapped Illumina paired-end reads of *Termitomyces* sp. J132 (Poulsen et al. 2014) (from which the homokaryon strain P5 was derived through protoplasting) (Nobre et al. 2014) to the mitochondrial genome of T132 (see below for details) to identify SNPs that could be used as strain-specific markers. We selected two markers in intergenic regions (MT3 and MT4), and one marker in a type 1 mobile intron (rec2) (Supplementary Table 1). The intergenic markers were used to track inheritance of each parental mitotype and potentially detect recombination of mtDNA (independent of introns). The intronic marker was used to detect transfer of introns between parental mitotypes, which could indicate co-occurrence of parental mitotypes in the cytoplasm and potential recombination of mtDNA flanking mobile introns. We initially included a second intron marker, 3AB, but over the course of the experiment we found no sign of

recombination for this marker from which we concluded this intron was not mobile, and the marker was dropped from the final data set.

The relative locations of marker regions in the mtDNA of T132 are shown in Supplementary Table 1. In addition, to distinguish heterokaryons from the parental homokaryons, we used a strain-specific SNP marker in the nuclear ITS region.

To analyze recombination of mtDNA in wild populations of *T. cryptogamus* we compared whole-genome Illumina sequence data of four additional heterokaryon strains from the wild (Table 1) to the complete mitochondrial genome sequences of P5 and T132. We mapped the reads of each strain to the mitochondrial genome of T132 using BowTie2 (Langmead & Salzberg 2012) in the program Ugene v.36.0 (Okonechnikov et al. 2012) with default settings for paired-end reads. We selected 46 polymorphisms (mostly point mutations) between P5 and T132 across the length of the mitochondrial genome, and for each of the four other strains we scored per polymorphism whether they matched P5 or T132. All polymorphisms were outside the coding regions of the core mitochondrial protein-coding genes (*cox1-3, nad1-6, atp6, atp8-9, cob*), as these regions contained no sequence variation among the six strains.

Table 1: *Termitomyces cryptogamus* strains used for this study

Strain	SRA accession
Termitomyces sp. T132	SRX4910412
Termitomyces sp. J132 (P5)	SRX255527
Termitomyces sp. T8	SRX10474843
Termitomyces sp. T28	SRX10474844
Termitomyces sp. T60a	SRX10474845
Termitomyces sp. T61	SRX10474846

Mating

Homokaryotic cultures of T132 and P5 were inoculated together on agar plates containing MYA medium (20g malt extract, 15g agar, 2g yeast extract per 1L) at a variable distance from each other (1cm-5cm) and incubated at 30 degrees centigrade. After 1-2 months, depending on inoculation distance, spores were harvested from the central hyphal mass between inoculation points. We suspended spores in sterilized MQ water and inoculated spore suspensions on fresh MYA plates, using sterile glass beads to disperse individual spores. After incubating at 30 degrees centigrade for 7-10 days, we harvested individual single-spore colonies and extracted DNA by microwaving colony tissue at 450W for ten minutes and dissolving in 1ml sterile MQ water.

To check for potential horizontal transfer of mobile introns, we sampled one plate more extensively from both the fusion zone and the halfway points between the fusion zone and the inoculation points (five samples per region), approximately seven days after the fusion zone became visible. The distance between the fusion zone and the inoculation points was about four centimeters.

Amplification and sequencing

To check if harvested spores were from heterokaryons or homokaryons, we first amplified ITS marker regions and used the ITS2 primer to sequence the PCR products (Sanger sequencing was performed by Eurofins Germany). When the sequencing product showed a peak for both strain-specific SNPs, we considered the singe-spore culture to derive from a heterokaryon. We amplified and sequenced the three mitochondrial marker regions for each heterokaryon culture as well as several homokaryon cultures (obtained by the same method from mating zones) as controls.

Our PCR protocol was as follows: we added 1µl DNA extraction to a mixture consisting of 5µl GoTaq buffer, 2µl 25mM MgCL2, 1µl DNTP, 0.5µl forward primer, 0.5µl reverse primer, 0.1µl GoTaq polymerase, 13.4µl MQ water, and 0.1µl BSA. We amplified DNA using a Touchdown PCR with an annealing temperature starting at 60 degrees centigrade and lowering at 0.5 degrees each cycle for 30 cycles. Amplified PCR products were diluted 20x in MQ water before Sanger sequencing.

Results

Mating Experiment

Our final data set from the mating experiment contained 31 heterokaryotic and 4 homokaryotic samples from seventeen independent mating zones, with 1-5 samples analyzed per mating zone. Another 10 homokaryotic samples were obtained from the flanks (202 and 203) of one mating zone (201). Sequencing results of the three mitochondrial markers are shown for each sample in Table 2.

For the non-intronic markers MT3 and MT4, we found eleven mating zones contained heterokaryons matching the P5 variant, while two mating zones contained heterokaryons matching T132. A single heterokaryon showed heteroplasmy for the MT3 marker, which was the only indication of both heteroplasmy and recombination for these markers. Of the mating zones for which we sampled multiple single spores, none showed variation among samples for these two markers.

Table 2: Sequencing results for three mitochondrial markers (rec2, MT3 and MT4) and one nuclear marker (ITS2) of single spore cultures obtained from fusion zones between two compatible strains of *Termitomyces cryptogamus*. Sequenced alleles for each marker are represented by A, B, or AB, indicating whether they match parent A, B, or heterozygosity, respectively.

				Markers			
Parent A	**Parent B**	**Mating zone**	**Single spore isolate**	**nuDNA ITS2**	**mtDNA intron rec2R**	**mtDNA MT3R**	**mtDNA MT4F**
hoP5	hoMn132	1	1	AB	A	A	A
hoP5	hoMn132	1	2	AB	A	A	A
hoP5	hoMn132	2	1	AB	AB	A	A
hoP5	hoMn132	2	2	AB	A	A	A
hoP5	hoMn132	3	1	AB	A	A	A
hoP5	hoMn132	5	1	AB	B	B	B
hoP5	hoMn132	5	2	AB	B	B	B
hoP5	hoMn132	5	4	AB	B	B	B
hoP5	hoMn132	6	1	A	A	A	A
hoP5	hoMn132	7	1	AB	B	A	A
hoP5	hoMn132	7	2	AB	B	A	A
hoP5	hoMn132	7	3	AB	B	A	A
hoP5	hoMn132	7	4	AB	B	A	A
hoP5	hoMn132	8	1	AB	A	A	A
hoP5	hoMn132	9	1	AB	B	A	A
hoP5	hoMn132	9	3	AB	A	A	A
hoP5	hoMn132	10	1	AB	B	AB	A
hoP5	hoMn132	11	1	AB	B	B	B
hoP5	hoMn132	11	2	AB	B	B	B
hoP5	hoMn132	12	1	A	A	A	A
hoP5	hoMn132	12	2	AB	A	A	A
hoP5	hoMn132	13	2	A	A	A	A
hoP5	hoMn132	14	1	A	A	A	A
hoP5	hoMn132	15	1	AB	B	A	A
hoP5	hoMn132	15	2	AB	A	A	A
hoP5	hoMn132	16	1	AB	A	A	A
hoP5	hoMn132	16	2	AB	A	A	A
hoP5	hoMn132	18	1	AB	A	A	A
hoP5	hoMn132	18	2	AB	A	A	A

hoP5	hoMn132	201	1	AB	B	A	A
hoP5	hoMn132	201	2	AB	B	A	A
hoP5	hoMn132	201	3	AB	A	A	A
hoP5	hoMn132	201	4	AB	A	A	A
hoP5	hoMn132	201	5	AB	A	A	A
hoP5	hoMn132	202	1	A	A	A	A
hoP5	hoMn132	202	2	A	A	A	A
hoP5	hoMn132	202	3	A	A	A	A
hoP5	hoMn132	202	4	A	A	A	A
hoP5	hoMn132	202	5	A	A	A	A
hoP5	hoMn132	203	1	B	B	B	B
hoP5	hoMn132	203	2	B	B	B	B
hoP5	hoMn132	203	3	B	B	B	B
hoP5	hoMn132	203	4	B	B	B	B
hoP5	hoMn132	203	5	B	B	B	B

For the marker rec2, located in a mobile intron, we found that for 10 out of 31 heterokaryon samples, the rec2 variant matched a different parent than the non-intronic MT3 and MT4 markers, indicating recombination. In all these cases the rec2 variant matched T132 while the MT3 and MT4 markers matched P5. In addition, we observed variation for rec2 among samples obtained from the same mating zone. We observed heteroplasmy for rec2 in a single heterokaryon. In the control homokaryon strains, we observed no recombination of the rec2 marker. We also observed no recombination of rec2 in the homokaryon single-spore cultures obtained from the halfway points (202 and 203) between the fusion zone and inoculation points (Figure 1).

We found two cases of heteroplasmy, once for the intron marker rec2 and once for marker MT3, each in a different single-spore culture. To control for contamination following DNA isolation we tested these samples twice for these markers. We consider contamination prior to DNA isolation unlikely as the other markers for these samples showed no sign of heteroplasmy. In both cases, none of the other mitochondrial markers from the same sample showed heteroplasmy.

We tested the homokaryon samples from the mating zones that had not become heterokaryon as controls. They were all P5 strains based on the nuclear ITS strains and all showed the P5 variant for the non-intronic markers MT3 and MT4 as well as for the intron marker rec2.

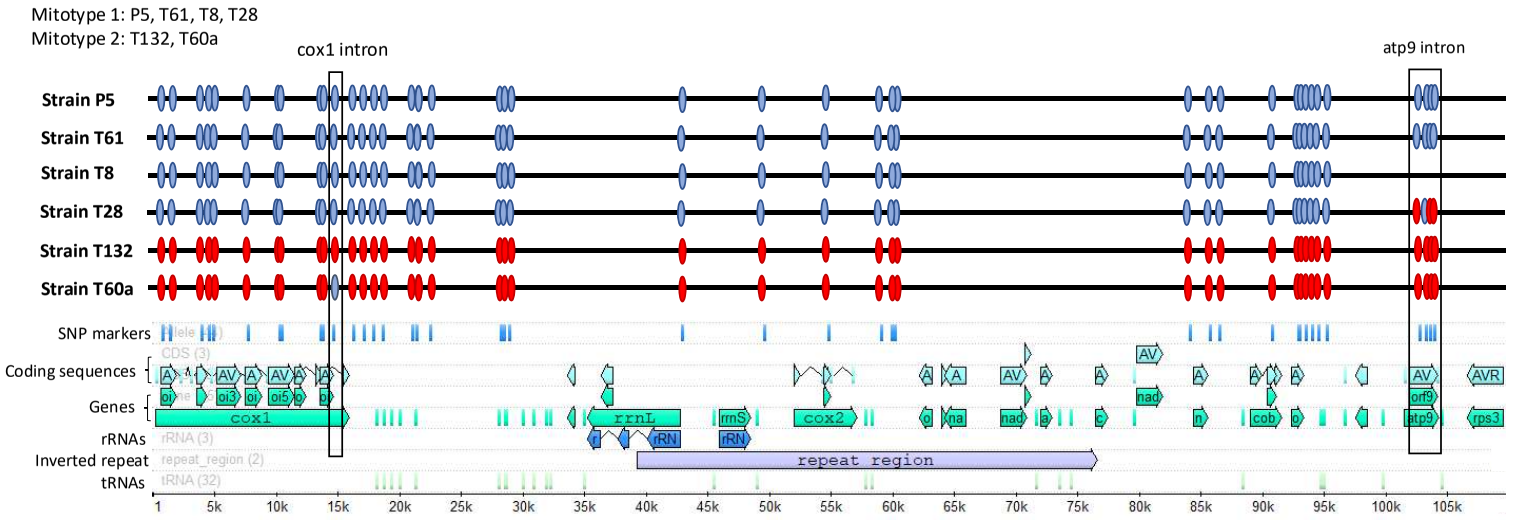


Figure 1: MtDNA recombination analysis of six wild strains of *Termitomyces cryptogamus*. We selected 46 SNP markers to test for recombinant mtDNA, which are shown for each strain in the upper bars at their corresponding position in the genome sequence of strain T132 (base pair positions shown on the bottom bar). Middle bars show the main annotated features of the mitochondrial genome of T132. Based on SNP markers, strains could be divided in two mitotypes (mitotype 1 or 2, color coded in top bars). Strains matched one mitotype for all markers except recombinant markers, which matched the other mitotype. Strain T60a had only one recombinant marker at base pair position 14518 in a cox1 intron. Strain T28 had three recombinant markers at positions 102603, 103500 and 103850 in the atp9 intron.

Population analysis

The results of the comparison of complete mitochondrial sequences of six wild strains of *T. cryptogamus* are shown in Figure 1. The six strains could be divided into two mitotypes (1 and 2) based on the selected SNP markers. P5, T61, T8 and T28 were considered to have mitotype 1, and T132 and T60a mitotype 2. Strains of each mitotype matched each other for all SNP markers except recombinant markers, in which case they matched the other mitotype. Strain T60a had a single recombinant marker located in the last intron of the *cox1* gene. Strain T28 had three recombinant markers located in the intron of *atp9*. Two of these were adjacent while the third was separated from the other two by a non-recombinant marker (Figure 1), indicating multiple recombination events within this intron.

Discussion

In this study we tested the mode of mitochondrial inheritance for a basidiomycete with limited nuclear migration, *Termitomyces cryptogamus,* and analysed the extent of mitochondrial recombination that occurs. Our experimental results, as well as the wild-population data indicate that mitochondrial inheritance in *Termitomyces* is effectively uniparental and lacks extensive recombination of mtDNA outside mobile introns.

Our results are inconclusive as to whether *Termitomyces cryptogamus* has limited or no nuclear migration following sexual reproduction. If nuclear migration is completely absent the heterokaryon must grow from the fused cells that are thought to be terminated in other basidiomycete matings (Rayner 1991; Ainsworth & Rayner 1986), in which the cytoplasms of both homokaryons are mixed. This would result in a transient phase of heteroplasmy during which recombination of mtDNA is possible. We observed only one case of mtDNA recombination outside mobile introns in our experimental mating study (Table 1, in a sample that was also heteroplasmic for one of the non-intron markers) and found no evidence in our wild population data. This suggests that if heteroplasmy occurs, it is generally short-lived and offers little to no chance for mitochondrial genomes to recombine. Since mitochondrial genomes are partitioned into nucleoids, which may not fuse frequently, recombination of different parental mtDNAs is probably rare in a short-lived heteroplasmic cell (Birky 2001). The question is whether recombination between mobile introns indicates a state of transient heteroplasmy. Since we did not observe intron recombination in homokaryon strains obtained from within the fusion zone or the flanking regions, it appears horizontal transfer independent of fusion during mating did not facilitate intron recombination in our experiment. Therefore, intron recombination suggests some heterokaryons have gone through a heteroplasmon stage either during mating or afterwards through fusion between resulting heterokaryons with different cytoplasm, and this heteroplasmon stage allowed mobile introns to recombine at a higher

frequency than 'normal' mtDNA recombination (with the latter possibly not occurring at all).

Another possibility is that in *T. cryptogamus* nuclei migrate a short distance through the homokaryon hyphae following mating. In this case the mycelial mass in the mating zone should be composed of two heterokaryons, one for each of the two parental mitotypes. Of the eleven mating zones for which we sampled multiple single spores, we found no case of heterogeneity for the non-intronic markers among single spore isolates (Table 1). We did find heterogeneity for the intron marker, however, which suggests that either there are multiple heterokaryons with recombined and non-recombined introns in the mating zone, or that single spores isolated from the fusion zone were heteroplasmic for the intron but became homoplasmic through continuous intron homing or drift during subsequent hyphal growth on plate.

In the wild, matings will often be asymmetric, for example between a newly germinated spore and an older mycelial monokaryon (Nieuwenhuis et al. 2013). In such a case, normal nuclear migration will result in the cytoplasm of the larger monokaryon dominating. However, if nuclear migration is limited the odds for each cytoplasm to succeed may be more even. In our experimental matings (Table 1), we observed the mtDNA of strain P5 more often in heterokaryon progeny than the mtDNA of strain T132. This could be an indication that, assuming limited nuclear migration results in roughly symmetrical mating, homoplasmic cells with P5 mtDNA occur more frequently than homoplasmic cells with T132 mtDNA. This could be the result of an increased growth or replication rate of cells with P5 mtDNA, and/or an increased chance of P5 mtDNA to outcompete T132 mtDNA in heteroplasmic cells. Alternatively, nuclear migration is not perfectly symmetrical, resulting in more cytoplasmic flow from P5 into T132.

We found two independent cases of heteroplasmy, once for marker rec2 and once for marker MT3 (Table 1). For marker rec2, it is possible that recombination was ongoing between the two intron types and homoplasmy had not yet been achieved as it had in other samples when DNA was extracted. It is more difficult to explain heteroplasmy for marker MT3, considering marker MT4 showed no sign of heteroplasmy. This suggests mtDNA recombination occurred outside of a mobile intron, which we did not observe in any other sample, and not in the wild samples either (Figure 1). Heteroplasmy for this marker suggests the recombinant mtDNA coexisted with non-recombinant mtDNA at the time of sampling.

Our results show that recombination of mtDNA in *Termitomyces cryptogamus* seems mostly restricted to homing endonucleases. Other studies on fungal mtDNA recombination should take endonuclease activity into account as they are ubiquitous elements in fungal

mtDNA and may be a major source for observed mitochondrial recombination events like those we report here in *Termitomyces*. To distinguish highly localized recombination due to homing endonucleases from general mtDNA recombination, it is important to include a large number of markers that provide a good coverage across the full length of the mitochondrial genome.

References

Ainsworth AM, Rayner ADM. 1986. Responses of Living Hyphae Associated with Self and Non-self Fusions in the Basidiomycete Phanerochaete velutina. Microbiology. 132:191–201. doi: https://doi.org/10.1099/00221287-132-1-191.

Anderson JB, Kohn LM. 2014. Dikaryons, Diploids, and Evolution. Sex Fungi. 333–348. doi: 10.1128/9781555815837.ch20.

Barr CM, Neiman M, Taylor DR. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytol. 168:39–50. doi: 10.1111/j.1469-8137.2005.01492.x.

Birky CW. 2001. The Inheritance of Genes in Mitochondria and Chloroplasts: Laws, Mechanisms, and Models. Annu. Rev. Genet. 35:125–148. doi: 10.1146/annurev.genet.35.102401.090231.

Christie JR, Schaerf TM, Beekman M. 2015. Selection against Heteroplasmy Explains the Evolution of Uniparental Inheritance of Mitochondria. PLoS Genet. 11:1–30. doi: 10.1371/journal.pgen.1005112.

Delahodde A et al. 1989. Site-specific DNA endonuclease and RNA maturase activities of two homologous intron-encoded proteins from yeast mitochondria. Cell. 56:431–441. doi: 10.1016/0092-8674(89)90246-8.

van Diepeningen AD et al. 2010. Mitochondrial recombination increases with age in Podospora anserina. Mech. Ageing Dev. 131:315–322. doi: https://doi.org/10.1016/j.mad.2010.03.001.

ELLIOTT TJ. 1978. Comparative Sexuality in Agaricus Species. Microbiology. 107:113–122. doi: https://doi.org/10.1099/00221287-107-1-113.

Fritsch ES, Chabbert CD, Klaus B, Steinmetz LM. 2014. A Genome-Wide Map of Mitochondrial DNA Recombination in Yeast. Genetics. 198:755–771. doi: 10.1534/genetics.114.166637.

Gogarten JP, Hilario E. 2006. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. BMC Evol. Biol. 6:94. doi: 10.1186/1471-2148-6-94.

Haugen P, Simon DM, Bhattacharya D. 2005. The natural history of group I introns. Trends Genet. 21:111–119. doi: https://doi.org/10.1016/j.tig.2004.12.007.

Hickey PC, Jacobson DJ, Read ND, Louise Glass N. 2002. Live-cell imaging of vegetative hyphal fusion in Neurospora crassa. Fungal Genet. Biol. 37:109–119. doi: https://doi.org/10.1016/S1087-1845(02)00035-X.

Hintz W, Anderson JB, Horgen PA. 1988. Nuclear migration and mitochondrial inheritance in the mushroom agaricus bitorquis. Genetics. 119:35–41. https://www.genetics.org/content/119/1/35.

Jin T, Sonnenberg ASM, Van Griensven LJLD, Horgen PA. 1992. Investigation of mitochondrial transmission in selected matings between homokaryons from commercial and wild-collected isolates of Agaricus bisporus (= Agaricus brunnescens). Appl. Environ. Microbiol. 58:3553–3560. doi: 10.1128/aem.58.11.3553-3560.1992.

De La Bastide PY, Horgen PA. 2003. Mitochondrial inheritance and the detection of non-parental mitochondrial DNA haplotypes in crosses of Agaricus bisporus homokaryons. Fungal Genet. Biol. 38:333–342. doi: 10.1016/S1087-1845(02)00584-4.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9:357–359. doi: 10.1038/nmeth.1923.

May G, Taylor JW. 1988. Patterns of mating and mitochondrial DNA inheritance in the agaric Basidiomycete Coprinus cinereus. Genetics. 118:213–220.

Mueller JE, Smith D, Belfort M. 1996. Exon coconversion biases accompanying intron homing: battle of the nucleases. Genes Dev. . 10:2158–2166. http://genesdev.cshlp.org/content/10/17/2158.abstract.

Nieuwenhuis BPS, Nieuwhof S, Aanen DK. 2013. On the asymmetry of mating in natural populations of the mushroom fungus Schizophyllum commune. Fungal Genet. Biol. 56:25–32. doi: 10.1016/j.fgb.2013.04.009.

Nieuwenhuis M et al. 2019. Enrichment of G4DNA and a Large Inverted Repeat Coincide in the Mitochondrial Genomes of Termitomyces. Genome Biol. Evol. 11. doi: 10.1093/gbe/evz122.

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. 14:1–12. doi: 10.1186/1471-2148-14-121.

Okonechnikov K et al. 2012. Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics. 28:1166–1167. doi: 10.1093/bioinformatics/bts091.

Poulsen M et al. 2014. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. Proc. Natl. Acad. Sci. U. S. A. 111:14500–14505. doi: 10.1073/pnas.1319718111.

Rayner ADM. 1991. The Challenge of the Individualistic Mycelium. Mycologia. 83:48–71. doi: 10.1080/00275514.1991.12025978.

Saville BJ, Kohli Y, Anderson JB. 1998. mtDNA recombination in a natural population. Proc. Natl. Acad. Sci. U. S. A. 95:1331–1335. doi: 10.1073/pnas.95.3.1331.

Wang P et al. 2017. Frequent heteroplasmy and recombination in the mitochondrial genomes of the basidiomycete mushroom Thelephora ganbajun. Sci. Rep. 7:1–14. doi: 10.1038/s41598-017-01823-z.

Wu B, Hao W. 2019. Mitochondrial-encoded endonucleases drive recombination of protein-coding genes in yeast. Environ. Microbiol. 21:4233–4240. doi: https://doi.org/10.1111/1462-2920.14783.

Xu J, Yan Z, Guo H. 2009. Divergence, hybridization, and recombination in the mitochondrial genome of the human pathogenic yeast Cryptococcus gattii. Mol. Ecol. 18:2628–2642. doi: 10.1111/j.1365-294X.2009.04227.x.

Xu J, Zhang Y, Pun N. 2013. Mitochondrial recombination in natural populations of the button mushroom Agaricus bisporus. Fungal Genet. Biol. 55:92–97. doi: 10.1016/j.fgb.2012.09.004.

Supplementary Figures

Supplementary Table 1: Primers used and their locations. Primer sequences in bold were used to sequence the amplified product.

Locus	Forward Primer	Reverse Primer	Location	Position
ITS1f-2	CTTGGTCATTTAGAGGAAGTAA	GCTGCGTTCTTCATCGATGC	Nuclear	31-1752
MT3	ACACTTCTGTCTCTCTGTC	**GAGCTAGGATACGAAGGTAT**	Mitochondrial, intergenic	86286 - 86557
MT4	**CTTCCAGATTCGAACTGGG**	CGTAGGGGTGCATTCGGATA	Mitochondrial, intergenic	99608 - 99950
rec2	TCCAAAGCCATTGCACTTAA	**ATCAGACTCCATGGCTATCA**	Mitochondrial, intron	54567- 54880
3AB	**CCACTCTGGACGGGATCAGT**	GGCTAGTATCGGGCTGATAA	Mitochondrial, intron	102337 - 102830

CHAPTER 6

6

General discussion

The papers in this thesis examine the origin and stability of mutualisms at various levels of biological complexity in *Termitomyces*: the domestication of *Termitomyces* by termites (Chapter 3), genetic addiction of mitochondria to plasmids (Chapter 4), and genetic variation of mitochondria after mating (Chapter 5). The mutualisms featured in each of these chapters occur at different tiers of biological complexity, and each has distinct characteristics and confounding factors complicating direct comparisons. However, they all are products of a similar evolutionary process that relegates individual performance in favor of group performance and must therefore address the problem of the tragedy of the commons. Here I will discuss how the differences and similarities between the mutualisms featured in this thesis inform us on fundamental questions regarding the origin and stability of cooperation in nature. After this section, I will focus on the specific topics of each chapter and discuss my findings for each in more depth.

The origin and stability of mutualisms

Mutualisms are unlikely to arise spontaneously between species that are not associated in some way beforehand. Therefore, commensals and parasites are prime candidates as initiators of a mutualism. Both the ancestor of *Termitomyces* and the ancestors of addictive mitochondrial plasmids were most likely a commensal of their host (termites and mitochondria, respectively), but it is also possible they were parasitic. In the case of the fungus-termite interaction, the formation of the mutualism probably involved a continuous, gradual process of co-evolution (Figure 1) similar to crop domestication in humans (Meyer & Purugganan 2013). Although this gradual process can proceed very rapidly, it involves a steady increase in co-dependency of both partners. By contrast, the mutualism between mitochondria and mitochondrial plasmids described in Chapter 4 arises abruptly following accidental loss of a tRNA-gene from the mitochondrial host. Rather than co-dependency evolving from an incremental process of co-evolution, dependency of the mtDNA on its plasmid is instantly imposed through a loss of function (Figure 1). Compared to its ancestor (prior to the loss of function) the mitochondrion has not necessarily gained a benefit from its association with the plasmid. This kind of mutualistic interaction is therefore perhaps more akin to a hostage situation rather than altruistic cooperation. Similar mutualisms arising by one partner holding the other hostage, for example by controlling their reproduction, are seen in selfish genetic elements (van der Gaag et al. 2000; Vogan et al. 2019; Courret et al. 2019), mitochondria (Horn et al. 2014; Chase 2007) and *Wolbachia* (Zug & Hammerstein 2015; Pannebakker et al. 2007; Zhang et al. 2020). These mutualisms are still expected on the long-term to develop into more benign variants from the host's perspective, as increased host fitness is in the interest of both partners. This transition is observed, among others, in *Wolbachia* (Zug & Hammerstein 2015; Zhang et al. 2020; Pannebakker et al. 2007).

A key question in evolutionary biology is the effect of genetic variation among symbionts in a host on the stability of a mutualism. A general assumption is that this genetic variation should be restricted by the host to avoid the emergence of selfish mutants that spread throughout the symbiont population. There is also some evidence for direct detrimental effects, for example of heteroplasmy on host cells (Sharpley et al. 2012). Examples of restriction of symbiont genetic variation include the uniparental transmission of organelles and the vertical transmission of *Buchnera* endosymbionts in pea aphids (Koga et al. 2012). Despite the obligate mutualism between *Termitomyces* and the Macrotermitinae, most termite hosts have not evolved a vertical transmission pathway for their symbionts, acquiring them horizontally instead by collecting new spores from their environment at the founding of each colony. If restricting genetic variation is important for the stability of a mutualism, how can the fungus-termite mutualism have prospered for over 30 million years?

The effect of symbiont genetic variation on the stability of a mutualism is likely dependent on the alignment of individual interests of host and symbiont and how these interact with the mutualism (Herre et al. 1999) (Figure 1). For example, what drives competition between divergent *Termitomyces* strains in the initial phase of comb construction by an incipient termite colony? Aanen et al. (Aanen et al. 2009) suggests that frequency-dependent selection by termite farmers would favor fast-growing, conidiospore-producing strains. How does this trait compare to the traits underpinning the fungus-termite mutualism? Although still uncertain, it is likely that *Termitomyces* provides an (indirect) nutritional service to the termites by digesting the comb material. Faster growth of the fungus may correlate with more efficient digestion of comb material (Lustenhouwer et al. 2020), and therefore the individual interests of host and symbiont in this mutualism could be aligned. Genetic variation and competition among symbionts then would not be a major concern for the host and could be beneficial, by selecting the fastest growers which are likely better digestive aids to the termite. Although competition between individuals also inhibits crop yield within a group (Denison et al. 2003), subsequent frequency-dependent selection by the termites establishes a monoculture of the dominant *Termitomyces* strain, resolving this issue long-term.

Let us contrast the fungus-termite mutualism to that between mitochondria and the eukaryote cell. The individual interest of a mitochondrion is reproduction: the more it replicates the more likely it is to dominate the cytoplasm and transfer to new cells. However, the mitochondrial trait that is important for the mutualism between mitochondria and the host cell is (among other services) the production of ATP. The production of ATP is not necessarily enhanced by faster replicating mitochondria. In fact, a well-documented example of selfish mtDNA is the 'petite' mutant in yeast (Bernardi 1979) that is completely respiratory-deficient and replicates much faster than its wildtype counterpart. The

individual interests of mtDNA and the eukaryote cell are not aligned, and competition among divergent mtDNAs can result in selfish mutants arising (Hastings 1992).

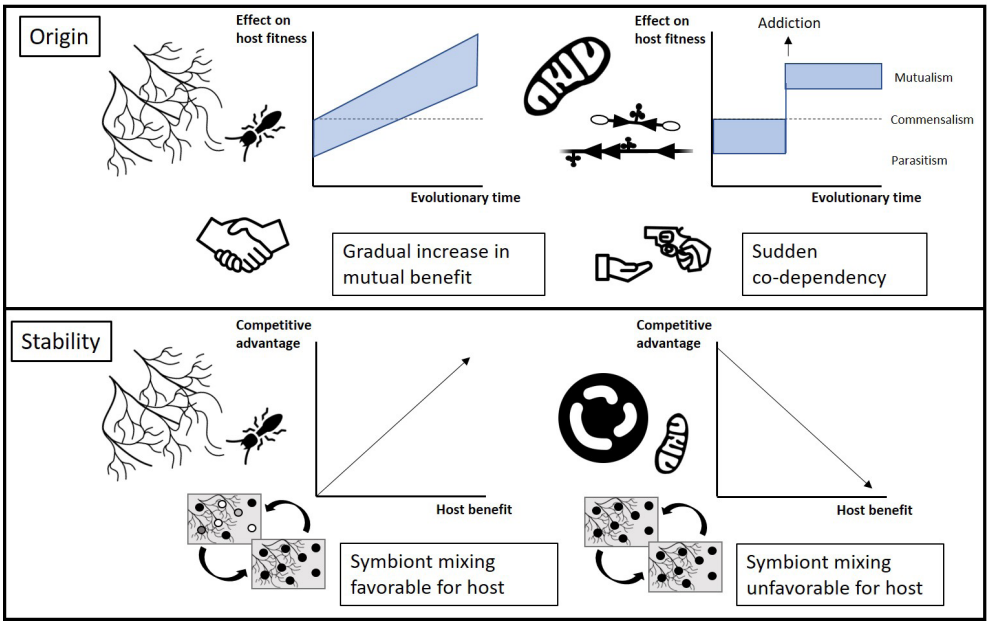


Figure 1: The origin and stability of mutualisms associated with *Termitomyces*

Mutualisms are likely to arise from pre-existing commensal or parasitic interactions. In the case of the fungus-termite mutualism, this likely involved a gradual process of co-adaptation between host and symbiont (top panel, left). By contrast, the mutualism between *Termitomyces* mtDNA and tRNA-carrying plasmids evolves abruptly by a loss-of-function on the part of the host, resulting in sudden addiction to the symbiont (top panel, right). Long-term stability of mutualisms depends on the potential for lower-level competition between symbionts to produce cheating individuals. In the fungus-termite mutualism, the competitive advantage of individual symbionts may be positively correlated to the benefit they provide to their host (bottom panel, left). As a result, genetic heterogeneity of symbionts may be beneficial to the host to generate more competitive and beneficial symbionts. The mutualism between eukaryote nuclei and mitochondria can be disrupted by accumulation of cheating mitochondria that invest in selfish replication at the expense of ATP production. In this situation the competitive advantage of individual mitochondria is negatively correlated to their benefit to the host (bottom panel, right). Since mixing different symbiont genotypes can be a detriment to the host, it is restricted through for example uniparental inheritance.

Evolution of the domestication of *Termitomyces*

In this thesis I have retraced the evolutionary history of *Termitomyces* and its closest relatives to attempt a reconstruction of the transition from free-living fungi to domesticated insect-symbionts. One hypothesis is that the ancestor of *Termitomyces* was a free-living wood-decomposing fungus that was brought into termite colonies as a 'stowaway' as termites collected dead plant matter as food (Mueller & Gerardo 2002). However, another possibility is that the ancestor of *Termitomyces* was a coprophilic fungus growing on insect feces, that adapted to growth on termite-built fecal combs.

To test which hypothesis is most likely, in Chapter 3 we reconstructed the phylogeny of *Termitomyces* and its closest known relatives in the Lyophyllaceae, and compared their ecologies and morphologies. We show that two close relatives of *Termitomyces* (*Arthromyces* and *Blastosporella*) are most likely coprophilic fungi that grow on insect feces. In addition, *Termitomyces*, *Arthromyces* and *Blastosporella* share a number of morphological traits, such as a rooting stipe and conidiospores, that facilitate a transition to the symbiotic lifestyle of *Termitomyces*. The hypothesis that the common ancestor of *Termitomyces* and *Arthromyces* grew on insect feces is supported by our finding that these two genera share a similar CAZyme complement, which indicates a similar diet. Together, these results suggest the ancestor of *Termitomyces* was coprophilic, already associated with insects and had a toolkit of traits that facilitated transition to life inside a termite mound.

If the ancestor of *Termitomyces* grew on insect feces, how did it end up in the interior of a termite mound? The sister clade of the fungus-farming Macrotermitinae comprises the Sphaerotermitinae (Bucek et al. 2019), which also make comb structures in their nests. However, these combs are colonized by bacteria, not fungi. If the ancestor of the Macrotermitinae made combs composed of vegetation and feces, these would have been a prime substrate for the ancestor of *Termitomyces*. The termites may have introduced fungal spores from the environment when constructing the combs. Over time, the ancestor of *Termitomyces* could have replaced any bacterial symbionts such as those that are still associated with the Sphaerotermitinae.

Our results indicate that *Termitomyces* was ancestrally associated with insects before its domestication by termites, which raises the question how other insect-fungus domestications originated. The attine ants have domesticated various fungi several times in multiple independent evolutionary events, in contrast to the single origin of *Termitomyces'* domestication (Mueller et al. 1998; Schultz & Brady 2008; Aanen et al. 2002). While the fungus-farming termites and *Termitomyces* both appear to be highly specialized towards their mutualistic lifestyle, the fungal crops of the attine ants show more variation

in the degree to which they co-evolved with their hosts (Nobre et al. 2011; Aanen et al. 2002; Villesen et al. 2004; De Fine Licht et al. 2014). Although the fungi associated with attine ants are unlikely to derive from coprophilic ancestors, it is possible they were ancestrally associated with insects for dispersal or as residents of ant nests built from leaf-litter (Mueller et al. 2001). For ambrosia beetles, some of their fungal symbionts are closely related to tree pathogenic fungi, which suggests their mutualism arose through a more coincidental, passive interaction (Cassar & Blackwell 1996).

A key element in the domestication of *Termitomyces* is the production of asexual spores (conidiospores) on the hyphae. These serve as 'seeds' for the termites to sow across fungus combs and enhance local dispersal of *Termitomyces*. Conidiospore production is likely an advantageous trait for coprophilic fungi to rapidly colonize ephemeral substrates such as feces. Consistent with this, we observed similar conidiospore production in the coprophilic relatives of *Termitomyces, Arthromyces,* and *Blastosporella*. In *Arthromyces*, the dendritic shape of the conidia suggests they attach to mobile animals for their dispersal, but we lack evidence to support this hypothesis. The presumed adaptation of plants to short-range dispersal by humans was an important transition in their domestication (Fuller & Allaby 2018). Similarly, the local dispersal of conidiospores by animals (possibly insects) may have been a key adaptation in the ancestor of *Termitomyces* that facilitated its domestication by the Macrotermitinae.

Although the evidence we have collected strongly suggests an insect-associated lifestyle preceding the domestication of *Termitomyces*, there are still gaps in in our knowledge that need to be addressed to complete our picture of the ancestral habits of these fungi. For one, we do not know the exact interactions *Arthromyces* and *Blastosporella* have with insects in terms of spore dispersal. The dendritic shape of the asexual spores of *Arthromyces* suggest they attach to animals for transportation, but this is an untested hypothesis. We also do not know whether *Blastosporella* represents an independent evolution to a coprophilic lifestyle from the *Termitomyces-Arthromyces* ancestor, since in our observations it forms a clade with at least one species (*Tephrocybe rancida*) that does not appear to be insect-associated. We are lacking data on the substrate and ecology of several other species in the termitomycetoid clade, including the sister clade to *Termitomyces-Arthromyces*, which consist of *Tephrocybe* sp. 4 and *Tricholoma furcatifolium*. Data on the niches of these species can help reconstruct the development of insect-associations in the termitomycetoid clade beyond the common ancestor of *Termitomyces* and *Arthromyces*.

Mitochondrial evolution in *Termitomyces* and other fungi

During this thesis project I sequenced and reconstructed whole mitochondrial genomes from a diverse selection of species from the Lyophyllaceae. This led to a number of unexpected discoveries regarding the structure and content of these genomes. I found mitochondrial genomes containing large, inverted repeats similar to those that partition chloroplast genomes in many plants (Chapter 2). I also discovered many of these mitochondrial genomes had a high G-quadruplex DNA content. Finally, I identified a large number of mitochondrial plasmids, including plasmid inserts in fungal mtDNA, as well as transfer of fungal mtDNA to plasmids (Chapter 4). To understand the evolutionary history and possible functional relevance of these features, it is important to analyse them in a broader, comparative context. In the following section I will discuss each of these three features (the IR, G4DNA, and plasmids) and link them to ongoing research in other organisms.

Inverted repeats

I discovered large, inverted repeats (IRs) in the mitochondrial genomes of *Termitomyces* and its sister species *Arthromyces* (unpublished data), *Blastosporella zonata,* and *Tephrocybe rancida* (Chapter 2). In some species, the combined length of both repeat copies covers more than half the mitochondrial genome. IRs of such size are rare in fungal mtDNA, but common in Oomycete mtDNA and plant chloroplast genomes. Their function if they have any, is unknown. A striking feature of these repeats is their perfect sequence similarity despite their often-considerable size. This preservation of each other's mirror image is due to continuous homologous recombination between the two pairs (Palmer 1983; Aldrich et al. 1985). When a mutation or sequence rearrangement occurs in one copy, homologous recombination with the other could either transfer the mutation to the other copy or restore the original sequence in the mutated repeat. Some studies suggest that the increased efficiency of homologous recombination-driven repair in IRs could be beneficial for repairing double-stranded breaks (Zhu et al. 2016; Gualberto et al. 2014), thus promoting genomic stability. In some plastid genomes, the mutation rate seems to be reduced within the IR (Palmer & Thompson 1982; Maier et al. 1995; Wu & Chaw 2015; Wolfe et al. 1987), which suggests there may be a bias towards restoring mutated repeat copies to their previous state, but this bias is not seen in other studies (Blazier et al. 2016), including our *Termitomyces* mtDNA analysis (Chapter 2). This discrepancy may indicate the relative frequency at which these organelle genomes occur in fused, concatameric states. If multiple copies of mtDNA are connected, there are many IR copies available for homologous recombination. A mutation in a single copy in that situation has a relatively high chance of being 'corrected' by an unmutated copy. However, if the mtDNA consists of a single 'monoploid' copy, the odds of a mutated IR copy being repaired in this way are presumably 50/50.

Another feature of large IRs is the instability of their flanking regions. In plant chloroplasts, the borders between the IR and single copy (SC) regions are known to shift constantly between species (Goulding et al. 1996). The IRs are constantly contracting and expanding, absorbing and expelling nearby sequences, including (partial) genes, which can result in rapid gene rearrangements as I observed in *Termitomyces*. Uptake of a gene flanking the 5′ side of an IR can later result in expulsion, by chance, of that gene on the 3′ side of the opposing copy, as I illustrate in Chapter 2, Figure 2. I found a peculiar discrepancy between the two flanks of the mitochondrial IRs in *Termitomyces* and its relatives. The flanks surrounding the SC1 region, which in all but *B. zonata* contained the cytochrome-oxidase 1 gene, appear very dynamic with frequent transposition of genes through contraction and expulsion of the IR (Chapter 2, Table 2). However, the flanks surrounding the SC2 region are by contrast very stable, always bordering cytochrome-oxidase 3 and the ribosomal protein S3 genes. Perhaps the SC1-IR border region is more prone to double-stranded breaks, or in some other way more frequently subjected to homologous-recombination, which in turn increases the chance of an erroneous non-homologous recombination event that can expand or contract the IR. Another possibility is that gene rearrangements in the SC2-IR border region are under strong negative selection.

The origin of the IR in the mtDNA of *Termitomyces* and its relatives remains unknown. In chloroplasts, the IRs can be traced back to the genome of the prokaryote ancestor (Palmer 1985; Turmel 1999), but in fungal mtDNA they have evolved independently in several lineages (Férandon et al. 2013; Fricova et al. 2010; Ferandon et al. 2008). In the fungus *Agrocybe aegerita*, a relatively small mitochondrial IR (~4kb) enclosed a short single copy region that was shown to be the inserted remnant of a mitochondrial plasmid (Ferandon et al. 2008). These plasmids, which are discussed more in depth in Chapter 4 and later in this section, are linear molecules capped by terminal inverted repeats in their free-living state. When inserting themselves in mtDNA, they tend to expand their IRs slightly with the mitochondrial sequences flanking the insertion site. From there they can potentially increase in size through non-homologous recombination errors, and this seems to be the case in *Agrocybe aegerita*. A recent study found a much larger IR in another species of *Agrocybe* (Liu et al. 2020), and although its origin is unknown, it could have derived from the same plasmid insertion. Since these plasmids and their mitochondrial inserts are abundant in *Termitomyces*, it is conceivable that the IR here also originated from a plasmid insertion.

G-quadruplex DNA

I discovered a high frequency of DNA motifs that are likely capable of G-quadruplex DNA formation in the mtDNA of most *Termitomyces* strains, as well as the other Lyophyllaceae with mitochondrial IRs (Chapter 2). Although G-quadruplex DNA (or G4DNA) formation has long been known to occur under natural conditions (Henderson et al. 1987), its effects

and potential functions are still mostly unknown and are probably dependent on the specific motif sequence and its location in the genome (Varshney et al. 2020). G4DNA sequences affect the genome by forming non-canonical structures that interrupt the 'standard' double helix conformation of DNA. Examples of locations rich in G4DNA are vertebrate telomeres and mtDNA. G4DNA is also often found near promotor sites of genes, including oncogenes (Varshney et al. 2020).

Since the functions of G4DNA are still unclear, genome-wide analyses of G4DNA distribution and its variation between species can provide insight in their evolution and potential effects. To date, G4DNA content has been studied in the genomes of humans, plants, and the fungus *Saccharomyces cerevisiae (*Capra et al. 2010; Chambers et al. 2015; Hershman et al. 2008; Garg et al. 2016). In Chapter 2, I provide the first comparative analysis of mitochondrial G4DNA content in fungi, showing it to vary significantly across a broad selection of species. I reported two observations that suggest the effects G4DNA may have on fungal mtDNA: 1) in species with mitochondrial IRs, I found on average more G4DNA motifs within the IR than outside, and 2) mitochondrial G4DNA in fungi is significantly reduced in coding regions compared to introns and intergenic sequences. I will discuss the significance of each of these findings separately.

The relative increase of G4DNA content in IR regions compared to SC regions as found in *Termitomyces* and its relatives could indicate a compensatory effect of the IR on double-stranded breaks (DSBs) caused by G4DNA formation. Assuming frequent DSBs are deleterious and subject to negative selection (Bharti, Joshua A Sommers, et al. 2014), I expect this selection to be weaker in the IRs due to its enhanced efficiency of DSB repair through homologous recombination-driven repair. This difference in the strength of negative selection on G4DNA formation between the IR and SC region could explain the relative difference in G4DNA content. By the same reasoning, if expansion of the IR results in better coverage of G4DNA, this expansion may be subject to positive selection. The strongest case of a correlation between G4DNA content and IR span is the mtDNA of *Termitomyces* sp. DKA19, in which the IR has contracted significantly compared to other *Termitomyces* species. The relatively meagre 4.4 kilobases of each IR copy are centred on a local cluster of G4DNA motifs, which combined for both copies contain 46% of the total G4DNA content of the mitochondrial genome, while the IR copies combined covered only 11% of the total mtDNA size. Given this skewness in mitochondrial G4DNA content in *T.* sp. DKA19, the relative position of its small IR seems correlated to this clustering. To better support these findings, it would be valuable to have data from evolutionarily independent cases of high genomic G4DNA content combined with large IRs. This could reveal whether the relative enrichment of IR regions with G4DNA is a general phenomenon or more specific to this clade of Lyophyllaceae.

To test the hypotheses on the role of the IR and the effects of G4DNA in the mtDNA of *Termitomyces*, we need more specific sequencing strategies to identify DSB and G4 formation *in situ*. Studying DSB frequencies in specific genomic locations, for example using quantitative DSB sequencing (Zhu et al. 2019), can answer the question whether they occur frequently near G4DNA motifs, and whether they occur more often within or outside the IR. In addition, the actual formation of G4DNA complexes can be analyzed by sequencing whole-genome amplicons with a G4 ligand (Yoshida et al. 2018). Using such techniques could reveal how many of the predicted G4DNA sequences in our study actually result in G4DNA structures and could thereby disrupt or regulate for example gene expression, and cause DSBs.

The second significant result was the observation that mitochondrial coding regions in fungi are significantly depleted in G4DNA content compared to non-coding regions. This is more striking when we compare the relative distribution of G4DNA motifs in human mtDNA. The mtDNA of *Homo sapiens* has an overall G4DNA content comparable to the upper range found in our analysis of fungal mtDNA (~1 motif per kb). However, in contrast to fungi, human mtDNA contains no introns and almost no intergenic DNA. Therefore, the G4DNA is almost exclusively concentrated in coding regions. I compared the expected to the observed distribution of G4DNA in the mitochondrial exons, introns, and intergenic DNA of *Termitomyces*, its relatives, and the ascomycete *Neurospora crassa* (Chapter 2, Figure 4). The G4DNA content of exons was significantly reduced in fungal mtDNA compared to the mtDNA of *Homo sapiens*. This suggests strong negative selection against G4DNA in mitochondrial coding regions in fungi compared to humans (and likely other vertebrates considering the similarity among vertebrate mtDNA). A recent study (Pietras et al. 2018) found that a combination of the mitochondrial degradation pathway and an RNA-mediating protein only found in vertebrates can selectively degrade G4DNA-prone transcripts. This is probably one mechanism by which vertebrates resolve the negative effects of G4DNA in coding sequences. It would be interesting to learn why this (potentially costly) mechanism evolved in vertebrates whereas in fungi G4DNA accumulation in coding regions is controlled by negative selection.

In summary, my comparative analysis of mitochondrial G4DNA content in fungi has generated new questions regarding the effects of G4DNA in genomes in general. I suggest the potentially negative impact of DSB formation by G4DNA can be compensated by homologous recombination-driven repair, which may have led to expansion of the mitochondrial IR in *Termitomyces* and its relatives. Furthermore, I propose that there is strong negative selection against G4DNA in mitochondrial exons in fungi, in contrast to vertebrates, which implies a fundamental difference between fungi and vertebrates in the consequences (and handling) of G4DNA formation in coding regions.

Linear mitochondrial plasmids

In Chapter 4, I describe a large number of linear mitochondrial plasmids found in samples of *Termitomyces* and other Lyophyllaceae. This type of plasmid is abundantly dispersed across the fungal kingdom, a result of millions, perhaps even billions of years of horizontal transfer and differential loss.

My principal finding was several independent transfers of mitochondrial tRNA genes to these autonomous plasmids, to my knowledge the first reports for this linear type. In two independent cases, this tRNA transfer coincided with loss of the gene in the mtDNA. Considering the codons coded for by these tRNAs were still found in the coding sequence of mitochondrial genes of the host, as well as the fact I never observed loss of mitochondrial tRNA genes in hosts in the absence of a plasmid-borne copy, I concluded that the host mitochondrion depends on the plasmid for its supply of this tRNA. This potentially represents a case of genetic addiction (Kobayashi 2004). An interesting dynamic to this addiction is the known habit of mitochondrial plasmids to frequently insert themselves into the host mtDNA. This gives the mitochondria a chance to recover the lost tRNA following plasmid insertion, possibly making the dependency only temporary. However, as long as a particular tRNA gene is lacking in the mtDNA but located on a plasmid, that plasmid would be insured of vertical transmission. Since plasmids may otherwise frequently be lost during vertical transmission (Van Der Gaag et al. 1998), such an abrupt co-dependency can pose a great selective advantage for the plasmid.

I would have liked to further test this dependency by comparing the rate of loss of 'addictive' plasmids to that of non-addictive plasmids (i.e., plasmids not carrying a tRNA or carrying a tRNA that is still present in the host mtDNA). I attempted inducing plasmid loss in strains of *Termitomyces* sp. T132, which hosts both an addictive and a non-addictive plasmid, using mutagenic UV radiation. However, I was unable to observe plasmid loss following various levels of UV dosage, despite significant fungal spore death (unpublished data). The fact that plasmids and their host mitochondria occur in multitude in each cell probably makes their elimination by mutagenesis difficult. In nature, loss of plasmids may occur during sexual spore formation which is accompanied by mitochondrial bottlenecks.

I also found that tRNAs transferred to plasmids coded for relatively rare codons in mitochondrial genes. I hypothesize that this may be a result of stochasticity if mitochondrial tRNA genes are transcribed at relatively equal rates. In that case the tRNA transcripts of rare anti-codons would be relatively underutilized by mRNAs, and thus potentially be available for accidental inclusion by the plasmid during its replication. Further studying this phenomenon, for example by testing the relative abundance of mitochondrial tRNA transcripts with targeted transcriptome sequencing, could give insight into the still poorly understood mechanisms of (fungal) mitochondrial genome

transcription and linear plasmid replication. Furthermore, I found that mtDNA transfer to linear mitochondrial plasmids has occurred on multiple, independent occasions, similar to what has been found for the rarer circular plasmids. This suggests that the replication mechanisms of linear and circular plasmids may share key similarities such as the use of a tRNA-like structure for replication initiation (Chiang & Lambowitz 1997; Kennell et al. 1994). This is interesting, since it was suggested that the replication mechanisms of circular mitochondrial plasmids could illustrate the evolutionary transition from the RNA to the DNA world (Mohr et al. 2000; Wang & Lambowitz 1993; Kuiper & Lambowitz 1988). If this is the case, further analysis of plasmid replication (linear and circular) could provide a rare glimpse at the mechanisms behind the origin of life itself.

Cytoplasmic inheritance and recombination of mtDNA

In Chapter 5 I analyzed mitochondrial sequence data of *Termitomyces* strains for signs of biparental cytoplasmic inheritance and mitochondrial recombination. The hypothesis that biparental inheritance and recombination of mtDNA occur in *Termitomyces* is based on similarities in the mode of sexual reproduction between *Termitomyces* and *Agaricus*, a basidiomycete genus in which these phenomena have been reported (Xu et al. 2013; De La Bastide & Horgen 2003). In both genera heterokaryon formation during mating occurs in the central mass of fused hypha, indicating limited nuclear migration (Nobre et al. 2014; Hintz et al. 1988).

Heteroplasmy and frequent recombination of mtDNA are assumed to be detrimental for the host cell as competition between different mtDNAs could select for selfish mutants that are inferior symbionts for the host (Christie et al. 2015; Hastings 1992). Nonetheless mtDNA recombination has been reported in various fungi (Wang et al. 2017; van Diepeningen et al. 2010; Brankovics et al. 2017; Fritsch et al. 2014; Xu et al. 2009). However, an important source of recombinant mtDNA may be homing endonucleases, which recombine separately from 'normal' mtDNA through a specialized self-splicing mechanism (Xu et al. 2009; van Diepeningen et al. 2010; Wu & Hao 2019).

My results show that recombination of mtDNA in *Termitomyces* seems mostly restricted to these homing endonucleases. Exchange of mtDNA through this homing mechanism in heterokaryon spores but not homokaryon spores obtained from mating zones suggests brief cytoplasmic contact (heteroplasmy) between cells containing different mtDNA. However, this transient phase of heteroplasmy does not facilitate (frequent) recombination of mtDNA outside this homing activity. Therefore, the increase in genetic diversity of mtDNA resulting from heteroplasmy is minimal and restricted to mutations within or flanking active endonucleases.

Without recombination, mtDNA could be susceptible to Muller's ratchet (Muller 1932), the process that describes the continuous accumulation of mostly detrimental mutations. Even a very low rate of recombination is theoretically sufficient to escape this problem (Charlesworth et al. 1993; Pamilo et al. 1987). Small-scale recombination events such as those occurring through homing endonuclease activity may satisfy the need for recombination to escape Muller's ratchet. At the same time, the restricted extent of this form of recombination may reduce the chance of selfish mtDNA mutants arising out of the moderate increase in genetic diversity.

Phylogenomics

Underpinning most of the work in this thesis is a data set of whole genome sequences, that I use both to describe characteristics of (mitochondrial) genomes for *Termitomyces* and several closely related fungi, as well as to reconstruct their evolutionary history (phylogenomics). As relatively young but rapidly advancing fields, genomics and phylogenomics offer an ever-increasing potential to study evolution at an unprecedented scale. However, they also present new challenges particularly in terms of handling, filtering, and interpreting the quantity of data genomic research generates. In the following section I briefly discuss recent developments in genome research and focus on some of the challenges relevant to this thesis.

The ability to obtain the complete genome sequence of an organism has opened a new frontier of exploration of the natural world. Whole-genome profiles are an invaluable resource to study evolutionary relationships, connect genetic variation to phenotypic traits for agricultural, industrial or medical purposes, and to study the function and properties of DNA itself. Rapid technological advances have increased the potency of genome sequencing and reduced its cost, allowing for an exponentially increasing record of genome assemblies covering more and more species from all major groups of life. However, this fast surge in quantity of genomic data has exposed bottlenecks in the processing and analysis of genome sequences. Annotation is difficult and often reliant on reference-based automated pipelines, which are custom-built for specific groups of organisms, not standardized, and that produce a variety of errors (Salzberg 2019; Ejigu & Jung 2020). Submission of genome data to public databases such as GenBank is time-consuming and frustrated by poor documentation, confusing user interfaces, and outdated software (Smith 2020; Benson et al. 2009). Certain analytical or statistical approaches, including phylogenetic bootstrapping, can poorly handle the volume of data contained in a genome sequence, and need to be adapted or replaced by different procedures (Shen et al. 2017; Felsenstein 1985; Minh et al. 2020). Finally, as we still know very little about many features of the genome, such as the role of non-coding DNA and

alternative structural conformations, much of the information contained in genome sequences is still hidden from us.

In Chapter 3, I annotated draft genome assemblies constructed using paired-end short read data. Annotating fragmented draft genomes is problematic because genes may be spread across different contigs, rendering them difficult to recognize (Salzberg 2019). One method to predict the severity of this problem for a particular genome assembly is to annotate a set of highly conserved genes that are expected to be present in the genome of that organism and see whether a significant percentage of those genes is missing from the annotation results. For our phylogenomic analysis, I annotated genes from a reference set of highly conserved basidiomycete genes (BUSCO) (Simão et al. 2015). BUSCO gene sets can be used to assess the completeness and quality of a genome assembly. From the annotation results I found most of our assemblies contained ~90% of BUSCO genes (unfragmented) indicating overall high quality of assemblies. For five assemblies however, the annotation results recovered less than 50% of BUSCO genes (unfragmented). Since it is unlikely that these species have lost a significant portion of such highly conserved genes, these assemblies are probably too fragmented or incomplete to retrieve the remaining BUSCO genes through annotation. We can assume that further annotation of other genes using the Funannotate pipeline (Palmer & Stajich 2019) will produce equally incomplete, and therefore unreliable, results for assemblies with poor BUSCO scores. For this reason, I excluded these assemblies from the CAZyme analyses, and although I have stored a sequence record of the assemblies on GenBank, I did not include annotations.

Felsenstein in his seminal paper on using bootstrap support to infer statistical significance of phylogenetic relationships predicted that his method would suffer from inflated false positive results if the number of characters used for bootstrapping becomes very high (Felsenstein 1985). This same effect is seen in posterior probability support values used in Bayesian phylogenetics (Shen et al. 2017). In Chapter 3, our phylogenomic data set consisted of 1131 genes, comprising well over a million characters. A preliminary maximum-likelihood tree produced using IQ-TREE1 (Nguyen et al. 2015) with standard ultrafast bootstrapping (Hoang et al. 2018) yielded 100% bootstrap support for all nodes, suggesting inflation. For our definitive analysis I used two new features of IQ-TREE2 to account for this inflation: 1) I used an adapted version of ultrafast bootstrapping that resamples partitions prior to resampling sites to reduce inflation (Gadagkar et al. 2005), and 2) I calculated gene and site concordance factors (Minh et al. 2020) to reveal conflict among gene trees and among sites for each node.

Using the adapted ultrafast bootstrapping method produced bootstrap values <100 for two nodes (BS:99, which still indicates significant support, and BS:72, which indicates insignificant support and a polytomy). Although I cannot say for certain that this reduction

in bootstrap support is due to the alternate bootstrapping approach, since I made several other adjustments between the trial and definitive phylogeny, these values show that the quantity of characters was not sufficient to saturate bootstrap values for at least some nodes for this analysis.

Gene and site concordance factors are a measure of conflict among gene trees and sites for a given node, and therefore add extra context to bootstrap values, which only indicate robustness of the node to changes in the underlying data. However, gene and site concordance factors, in contrast to bootstrap values, have no statistical significance and therefore cannot be used to objectively express confidence in nodes. Rather, they provide an intuitive notion of conflict among gene trees or sites for a given node.

Coalescence versus concatenation

The vast quantity of genomic data available for phylogenetic analysis in recent times has heated a debate on how to correctly infer species phylogenies from genomic data (Bravo et al. 2019). Considering genes can evolve independently from each other for various reasons such as recombination, hybridization, incomplete lineage sorting, and horizontal transfer, an important problem in phylogenetics is how to infer the correct species tree from conflicting gene trees using a bifurcating pattern. One approach is to sum up all state changes from all genes (creating a single concatenated 'supergene') and assuming the phylogeny reconstructed from this supergene is the best reflection of the species tree. This concatenation approach presupposes that the average phylogenetic signal from all genes combined drowns out any internal conflicts between gene trees. However, simulation studies have shown that increasing data volume in a concatenation analysis can actually produce statistically inconsistent results and inflate the likelihood support for incorrect topologies (Kubatko & Degnan 2007). Alternatively, one could infer each gene tree separately and apply a greedy consensus method to reconstruct the species tree based on the most common gene tree topology. A problem with this approach is that it is possible that the most frequent gene tree topology is incongruent with the actual species tree. This can happen when the species tree includes more than four taxa, and the gene tree is in the so-called 'anomaly zone' of parameter space (Degnan & Rosenberg 2006).

Another approach is to apply a tree building algorithm that uses the multi-species coalescent model, such as ASTRAL (Mirarab et al. 2014). These methods generally break down gene trees into unrooted, four-leaved trees called quartets, one for each node. Since quartets cannot be in the aforementioned anomaly zone, they are more appropriate to reconstruct a greedy consensus tree than full gene trees. ASTRAL can account for gene tree conflict due to incomplete lineage sorting, but not gene tree conflict caused by other phenomena such as recombination (Mirarab et al. 2014).

In addition to the IQ-TREE concordance analysis, I performed an ASTRAL analysis (Mirarab et al. 2014) on our set of gene trees to see whether gene-tree incongruence due to incomplete lineage sorting affected the species-tree reconstruction. I found the ASTRAL tree topology was congruent with the concatenation-based IQ-TREE phylogeny. This suggests that incomplete lineage sorting was not a significant factor affecting the IQ-TREE analysis. Combining concatenation and coalescent approaches in a phylogenomic study increases reliability of the results in a similar way to combined maximum-likelihood and Bayesian approaches, by showing whether results are consistent under different sets of assumptions. Both concatenation and coalescence-based methods can produce erroneous results when their underlying assumptions, such as no recombination, are violated. The debate on which method is more appropriate for phylogenomic studies is ongoing (Liu et al. 2019), and many studies include both to account for potential conflicts.

Final conclusions and future prospects

In this thesis I have studied and compared aspects of three different mutualisms in the basidiomycete *Termitomyces*: the origin and evolution of the fungus-termite mutualism (Chapter 3), the addiction of mtDNA to plasmids (Chapter 4), and the inheritance and recombination of mtDNA (Chapter 5). In addition, the reconstruction and analysis of mitochondrial genomes for several *Termitomyces* and related species (Chapter 2) revealed new information on the properties and variation of fungal mtDNA and associated G4DNA motifs.

I set out to gain more understanding on how mutualisms arise and persist despite the risk of disruptive lower-level selection for cheaters. In the case of the fungus-termite mutualism, we found evidence for a pre-existing, possibly commensal interaction between the ancestor of *Termitomyces* and insects. From such an association the gradual development of a mutualism through incremental service exchange seems plausible. The potential for cheating by the fungal symbiont in this mutualism may be limited because the individual competitive benefit of the fungi may be correlated with their beneficial service to the host termites (Figure 1), circumventing the problem illustrated by the parable of the tragedy of the commons.

The mutualism between the mtDNA of some *Termitomyces* species and tRNA-carrying mitochondrial plasmids (Chapter 4) is an example of a pre-existing neutral or even parasitic association abruptly turning mutualistic through a loss of function on the part of the host; this function is then taken over by the symbiont. The long-term success of these mutualistic plasmids may not be very high due to their habit of reintegrating into the host mtDNA and thereby restoring the adopted function. However, they illustrate a general mechanism by which mutualistic associations can arise very rapidly: by symbionts taking over an essential function of their host.

The inheritance of mitochondria in *Termitomyces* appears to proceed with limited effects of heteroplasmy and recombination (Chapter 5). Although my results could not rule out biparental inheritance and cytoplasmic mixing during sexual reproduction, they showed no sign of frequent mtDNA recombination or heteroplasmy in both experimental crossings and population-level mtDNA sequence comparison. I observed limited recombination near homing endonucleases that are known to recombine separately from other mtDNA regions. These results are in line with an expected restriction of genetic divergence of mtDNA within the host cytoplasm, possibly due to a vulnerability of the nuclear-mitochondrial mutualism to cheating (Figure 1).

This thesis makes a small contribution to the ongoing research on the evolution of cooperation and the major transitions in life's complexity. By itself it does not answer any fundamental questions regarding the origin and stability of mutualisms, but it adds to an accumulating body of evidence that can be used to address such questions. Of equal importance, this thesis has produced a wealth of genomic information for the genus *Termitomyces* and many other species of the Lyophyllaceae, a diverse and widely studied group of fungi. This data facilitates new research and the results of this thesis give indications to promising new directions for these studies.

Regarding the ancestral condition of *Termitomyces,* we now have strong evidence that a pre-existing insect association facilitated the transition to domestication by termites. However, due to uncertainty and missing information regarding the ecology and habits of related fungi like *Tricholoma furcatifolium* and several *Tephrocybe* species, we cannot yet fully reconstruct the origin of the insect association itself. In addition, we lack observational evidence regarding the dispersal of spores by *Arthromyces* and *Blastosporella*, which could tell us more about their dependency on insects beyond being providers of a growth substrate. Are the fungi also consumed by potential insect hosts, and could their conidiospores survive a gut passage? The answers to these questions can inform us on the specific order of steps the ancestors of *Termitomyces* took between a free-living and symbiotic lifestyle. Ancestral reconstruction of other domesticated fungi like the symbionts of the attine ants and ambrosia beetles can then be compared to that of *Termitomyces* to find common patterns between independent transitions from free-living to domesticated states. Most important to further research in this area will be additional field work to observe these fungi and their natural environment.

The discoveries I made in Chapter 2 regarding the high G4DNA content of the mtDNA of several fungi, including *Termitomyces* and *Neurospora*, open new questions regarding the effects and potential functions of G4DNA in these organelles. In particular, the fact that the fungal G4DNA motifs are biased towards non-coding regions is interesting, as it suggests a disruptive interaction between G4DNA and transcribed RNA. This negative effect appears to be mitigated or absent in humans as our mtDNA is equally dense in G4DNA while virtually lacking non-coding regions. Nonetheless the G4DNA in human mitochondria seems to be involved in gradual mutation accumulation and degeneration of mtDNA function (Bharti, Joshua A. Sommers, et al. 2014). Some *Neurospora* strains can be cultured indefinitely in the lab, and this effective immortality is perhaps in part due to the absence of G4DNA in mitochondrial coding regions. Experimental genomic studies to observe mitochondrial DNA integrity of *Neurospora* around G4DNA regions could inform us whether G4DNA compromises mtDNA stability in this fungus. The formation of G4DNA structures can be observed using spectral or nanopore techniques, among others, and

their effects on surrounding DNA integrity can be assessed using for example qPCR or whole-genome sequencing to look for traces of double-stranded breaks (Zhu et al. 2019).

My finding of potential genetic addiction of *Termitomyces* mtDNA to tRNA-carrying plasmids adds a new perspective to these abundant genetic symbionts of fungal mitochondria. Ideally, I would have experimentally verified the dependency of *Termitomyces* strains lacking a mitochondrial tRNA gene to their associated plasmid. Considering that vertical transmission is the most likely natural pathway through which plasmids are lost (Van Der Gaag et al. 1998), a good test would be to sequence a large number of basidiospores from different individuals of *Termitomyces* sp. T132. This species lacks a mitochondrial copy of a tRNA-Arg gene, while carrying two plasmid types, one of which encodes a copy of this tRNA gene. If this tRNA-carrying plasmid is always present in single-spore cultures obtained from these basidiospores, while the other plasmid is occasionally lost, this would be strong evidence that the tRNA-carrying plasmid is essential for its host.

In Chapter 5 I found that mitochondrial recombination in *Termitomyces* is restricted to small regions encoding homing endonucleases. The question remains to what extent heteroplasmy occurs and, if it does, how 'regular' recombination of mtDNA is restricted. Physical tracking of mitochondria during mating of *Termitomyces,* using for example fluorescent markers, can reveal how the mitochondria move and whether they mix or are partitioned in fused cells. Further study of cytoplasmic inheritance in *Termitomyces* can reveal more about the role of nuclear migration in preventing heteroplasmy during basidiomycete mating, the mobility of homing endonucleases and their impact on genetic variation of mtDNA, and the mechanisms by which fungi resolve heteroplasmy if it occurs.

References

Aanen DK et al. 2009. High symbiont relatedness stabilizes mutualistic cooperation in fungus-growing termites. Science. 326:1103–1106. doi: 10.1126/science.1173462.

Aanen DK et al. 2002. The evolution of fungus-growing termites and their mutualistic fungal symbionts. Proc. Natl. Acad. Sci. U. S. A. 99:14887–14892. doi: 10.1073/pnas.222313099.

Aldrich J, Cherney B, Merlin E, Williams C, Mets L. 1985. Recombination within the inverted repeat sequences of the Chlamydomonas reinhardii chloroplast genome produces two orientation isomers. Curr. Genet. 9:233–238. doi: 10.1007/BF00420317.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2009. GenBank. Nucleic Acids Res. 37:D26–D31. doi: 10.1093/nar/gkn723.

Bernardi G. 1979. The petite mutation in yeast. Trends Biochem. Sci. 4:197–201. doi: https://doi.org/10.1016/0968-0004(79)90079-3.

Bharti SK, Sommers Joshua A., 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. 289:29975–29993. doi: 10.1074/jbc.M114.567073.

Bharti SK, Sommers Joshua A, 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. 289:29975–29993. doi: 10.1074/jbc.M114.567073.

Blazier JC et al. 2016. Variable presence of the inverted repeat and plastome stability in Erodium. Ann. Bot. 117:1209–1220. doi: 10.1093/aob/mcw065.

Brankovics B et al. 2017. Mitochondrial genomes reveal recombination in the presumed asexual Fusarium oxysporum species complex. BMC Genomics. 18:1–14. doi: 10.1186/s12864-017-4116-5.

Bravo GA et al. 2019. Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. PeerJ. 7:e6399. doi: 10.7717/peerj.6399.

Bucek A et al. 2019. Evolution of Termite Symbiosis Informed by Transcriptome-Based Phylogenies. Curr. Biol. 29:3728-3734.e4. doi: https://doi.org/10.1016/j.cub.2019.08.076.

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. doi: 10.1371/journal.pcbi.1000861.

Cassar S, Blackwell M. 1996. Convergent origins of ambrosia fungi. Mycologia. 88:596–601.

Chambers VS et al. 2015. High-throughput sequencing of DNA G-quadruplex structures in the human genome. Nat. Biotechnol. 33:877–881.

Charlesworth D, Morgan MT, Charlesworth B. 1993. Mutation accumulation in finite outbreeding and inbreeding populations. Genet. Res. (Camb). 61:39–56.

Chase CD. 2007. Cytoplasmic male sterility: a window to the world of plant mitochondrial–nuclear interactions. Trends Genet. 23:81–90. doi: https://doi.org/10.1016/j.tig.2006.12.004.

Chiang CC, Lambowitz AM. 1997. The Mauriceville retroplasmid reverse transcriptase initiates cDNA synthesis de novo at the 3' end of tRNAs. Mol. Cell. Biol. 17:4526 LP – 4535. doi: 10.1128/MCB.17.8.4526.

Christie JR, Schaerf TM, Beekman M. 2015. Selection against Heteroplasmy Explains the Evolution of Uniparental Inheritance of Mitochondria. PLoS Genet. 11:1–30. doi: 10.1371/journal.pgen.1005112.

Courret C, Chang C-H, Wei KH-C, Montchamp-Moreau C, Larracuente AM. 2019. Meiotic drive mechanisms: lessons from *Drosophila*. Proc. R. Soc. B Biol. Sci. 286:20191430. doi: 10.1098/rspb.2019.1430.

Degnan JH, Rosenberg NA. 2006. Discordance of species trees with their most likely gene trees. PLoS Genet. 2:e68–e68. doi: 10.1371/journal.pgen.0020068.

Denison RF, Kiers ET, West SA. 2003. Darwinian Agriculture: When Can Humans Find Solutions Beyond the Reach of Natural Selection? Q. Rev. Biol. 78:145–168. doi: 10.1086/374951.

van Diepeningen AD et al. 2010. Mitochondrial recombination increases with age in Podospora anserina. Mech. Ageing Dev. 131:315–322. doi: https://doi.org/10.1016/j.mad.2010.03.001.

Ejigu GF, Jung J. 2020. Review on the Computational Genome Annotation of Sequences Obtained by Next-Generation Sequencing. Biology (Basel). 9:295. doi: 10.3390/biology9090295.

Felsenstein J. 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution (N. Y). 39:783–791. doi: 10.2307/2408678.

Ferandon C, Chatel SEK, Castandet B, Castroviejo M, Barroso G. 2008. The Agrocybe aegerita mitochondrial genome contains two inverted repeats of the nad4 gene arisen by duplication on both sides of a linear plasmid integration site. Fungal Genet. Biol. 45:292–301. doi: 10.1016/j.fgb.2007.10.012.

Férandon C, Xu J, Barroso G. 2013. The 135 kbp mitochondrial genome of Agaricus bisporus is the largest known eukaryotic reservoir of group I introns and plasmid-related sequences. Fungal Genet. Biol. 55:85–91. doi: 10.1016/j.fgb.2013.01.009.

De Fine Licht HH, Boomsma JJ, Tunlid A. 2014. Symbiotic adaptations in the fungal cultivar of leaf-cutting ants. Nat. Commun. 5:5675. doi: 10.1038/ncomms6675.

Fricova D et al. 2010. The mitochondrial genome of the pathogenic yeast Candida subhashii: GC-rich linear DNA with a protein covalently attached to the 5′ termini. Microbiology. 156:2153–2163. doi: 10.1099/mic.0.038646-0.

Fritsch ES, Chabbert CD, Klaus B, Steinmetz LM. 2014. A Genome-Wide Map of Mitochondrial DNA Recombination in Yeast. Genetics. 198:755–771. doi: 10.1534/genetics.114.166637.

Fuller DQ, Allaby R. 2018. Seed Dispersal and Crop Domestication: Shattering, Germination and Seasonality in Evolution Under Cultivation. Annu. Plant Rev. online. 238–295. doi: https://doi.org/10.1002/9781119312994.apr0414.

van der Gaag M et al. 2000. Spore-killing meiotic drive factors in a natural population of the fungus Podospora anserina. Genetics. 156:593–605.

Van Der Gaag M, Debets AJM, Osiewacz HD, Hoekstra RF. 1998. The dynamics of pAL2-1 homologous linear plasmids in Podospora anserina. Mol. Gen. Genet. 258:521–529. doi: 10.1007/s004380050763.

Gadagkar SR, Rosenberg MS, Kumar S. 2005. Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree. J. Exp. Zool. Part B Mol. Dev. Evol. 304B:64–74. doi: https://doi.org/10.1002/jez.b.21026.

Garg R, Aggarwal J, Thakkar B. 2016. Genome-wide discovery of G-quadruplex forming sequences and their functional relevance in plants. Sci. Rep. 6:28211. doi: 10.1038/srep28211.

Goulding SE, Wolfe KH, Olmstead RG, Morden CW. 1996. Ebb and flow of the chloroplast inverted repeat. Mol. Gen. Genet. MGG. 252:195–206. doi: 10.1007/BF02173220.

Gualberto JM et al. 2014. The plant mitochondrial genome: Dynamics and maintenance. Biochimie. 100:107–120. doi: 10.1016/j.biochi.2013.09.016.

Hastings IM. 1992. Population genetic aspects of deleterious cytoplasmic genomes and their effect on the evolution of sexual reproduction. Genet. Res. 59:215–225. doi: 10.1017/S0016672300030500.

Henderson E, Hardin CC, Walk SK, Tinoco I, Blackburn EH. 1987. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine·guanine base pairs. Cell. 51:899–908. doi: https://doi.org/10.1016/0092-8674(87)90577-0.

Herre EA, Knowlton N, Mueller UG, Rehner SA. 1999. The evolution of mutualisms: exploring the paths between conflict and cooperation. Trends Ecol. Evol. 14:49–53. doi: https://doi.org/10.1016/S0169-5347(98)01529-8.

Hershman SG et al. 2008. Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. Nucleic Acids Res. 36:144–156. doi: 10.1093/nar/gkm986.

Hintz W, Anderson JB, Horgen PA. 1988. Nuclear migration and mitochondrial inheritance in the mushroom agaricus bitorquis. Genetics. 119:35–41. https://www.genetics.org/content/119/1/35.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. 35:518–522. doi: 10.1093/molbev/msx281.

Horn R, Gupta KJ, Colombo N. 2014. Mitochondrion role in molecular basis of cytoplasmic male sterility. Mitochondrion. 19:198–205. doi: https://doi.org/10.1016/j.mito.2014.04.004.

Kennell JC, Wang H, Lambowitz AM. 1994. The Mauriceville plasmid of Neurospora spp. uses novel mechanisms for initiating reverse transcription in vivo. Mol. Cell. Biol. 14:3094–3107. doi: 10.1128/mcb.14.5.3094.

Kobayashi I. 2004. Genetic Addiction: a Principle of Gene Symbiosis in a Genome. Plasmid Biol. 105–144. doi: https://doi.org/10.1128/9781555817732.ch6.

Koga R, Meng X-Y, Tsuchida T, Fukatsu T. 2012. Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte{\textendash}embryo interface. Proc. Natl. Acad. Sci. 109:E1230--E1237. doi: 10.1073/pnas.1119212109.

Kubatko LS, Degnan JH. 2007. Inconsistency of Phylogenetic Estimates from Concatenated Data under Coalescence. Syst. Biol. 56:17–24. doi: 10.1080/10635150601146041.

Kuiper MTR, Lambowitz AM. 1988. A novel reverse transcriptase activity associated with mitochondrial plasmids of neurospora. Cell. 55:693–704. doi: 10.1016/0092-8674(88)90228-0.

De La Bastide PY, Horgen PA. 2003. Mitochondrial inheritance and the detection of non-parental mitochondrial DNA haplotypes in crosses of Agaricus bisporus homokaryons. Fungal Genet. Biol. 38:333–342. doi: 10.1016/S1087-1845(02)00584-4.

Liu L, Anderson C, Pearl D, Edwards S V. 2019. Modern Phylogenomics: Building Phylogenetic Trees Using the Multispecies Coalescent Model BT - Evolutionary Genomics: Statistical and Computational Methods. In: Anisimova, M, editor. Springer New York: New York, NY pp. 211–239. doi: 10.1007/978-1-4939-9074-0_7.

Liu X, Wu X, Tan H, Xie B, Deng Y. 2020. Large inverted repeats identified by intra-specific comparison of mitochondrial genomes provide insights into the evolution of Agrocybe aegerita. Comput. Struct. Biotechnol. J. 18:2424–2437. doi: https://doi.org/10.1016/j.csbj.2020.08.022.

Lustenhouwer N et al. 2020. A trait-based understanding of wood decomposition by fungi. Proc. Natl. Acad. Sci. 117:11551–11558. doi: 10.1073/pnas.1909166117.

Maier RM, Neckermann K, Igloi GL, Koössel 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. 251:614–628. doi: 10.1006/jmbi.1995.0460.

Meyer RS, Purugganan MD. 2013. Evolution of crop species: genetics of domestication and diversification. Nat. Rev. Genet. 14:840–852. doi: 10.1038/nrg3605.

Minh BQ, Hahn MW, Lanfear R. 2020. New Methods to Calculate Concordance Factors for Phylogenomic Datasets. Mol. Biol. Evol. 37:2727–2733. doi: 10.1093/molbev/msaa106.

Mirarab S et al. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics. 30:i541–i548. doi: 10.1093/bioinformatics/btu462.

Mohr S, Wanner LA, Bertrand H, Lambowitz AM. 2000. Characterization of an unusual tRNA-like sequence found inserted in a Neurospora retroplasmid. Nucleic Acids Res. 28:1514–1524. doi: 10.1093/nar/28.7.1514.

Mueller UG, Gerardo N. 2002. Fungus-farming insects: Multiple origins and diverse evolutionary histories. Proc. Natl. Acad. Sci. 99:15247–15249. doi: 10.1073/pnas.242594799.

Mueller UG, Rehner SA, Schultz TR. 1998. The Evolution of Agriculture in Ants. Science (80-.). 281:2034–2038. doi: 10.1126/science.281.5385.2034.

Mueller UG, Schultz TR, Currie CR, Malloch D. 2001. The Origin of the Attine Ant-Fungus Mutualism. Q. Rev. Biol. 76:169–197. doi: 10.1086/393867.

Muller HJ. 1932. Some Genetic Aspects of Sex. Am. Nat. 66:118–138. doi: 10.1086/280418.

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. 32:268–274. doi: 10.1093/molbev/msu300.

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. 14:1–12. doi: 10.1186/1471-2148-14-121.

Nobre T, Rouland-Lefèvre C, Aanen DK. 2011. Comparative Biology of Fungus Cultivation in Termites and Ants BT - Biology of Termites: a Modern Synthesis. In: Bignell, DE, Roisin, Y, & Lo, N, editors. Springer Netherlands: Dordrecht pp. 193–210. doi: 10.1007/978-90-481-3977-4_8.

Palmer J, Stajich J. 2019. nextgenusfs/funannotate: funannotate v1.5.3. doi: 10.5281/zenodo.2604804.

Palmer JD. 1983. Chloroplast DNA exists in two orientations. Nature. 301:92–93. doi: 10.1038/301092a0.

Palmer JD. 1985. Comparative organization of chloroplast genomes. Annu. Rev. Genet. 19:325–354. doi: 10.1146/annurev.ge.19.120185.001545.

Palmer JD, Thompson WF. 1982. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell. 29:537–550. doi: 10.1016/0092-8674(82)90170-2.

Pamilo P, Nei M, Li WH. 1987. Accumulation of mutations in sexual and asexual populations. Genet. Res. 49:135–146. doi: 10.1017/s0016672300026938.

Pannebakker BA, Loppin B, Elemans CPH, Humblot L, Vavre F. 2007. Parasitic inhibition of cell death facilitates symbiosis. Proc. Natl. Acad. Sci. 104:213–215. doi: 10.1073/pnas.0607845104.

Pietras Z et al. 2018. Dedicated surveillance mechanism controls G-quadruplex forming non-coding RNAs in human mitochondria. Nat. Commun. 9. doi: 10.1038/s41467-018-05007-9.

Salzberg SL. 2019. Next-generation genome annotation: we still struggle to get it right. Genome Biol. 20:92. doi: 10.1186/s13059-019-1715-2.

Schultz TR, Brady SG. 2008. Major evolutionary transitions in ant agriculture. Proc. Natl. Acad. Sci. 105:5435–5440. doi: 10.1073/pnas.0711024105.

Sharpley MS et al. 2012. Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and cognition. Cell. 151:333–343. doi: 10.1016/j.cell.2012.09.004.

Shen X-X, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat. Ecol. Evol. 1:126. doi: 10.1038/s41559-017-0126.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31:3210–3212. doi: 10.1093/bioinformatics/btv351.

Smith DR. 2020. Depositing annotated sequences in GenBank: there needs to be a better way. Brief. Funct. Genomics. 19:337–338. doi: 10.1093/bfgp/elaa016.

Turmel M. 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. 96:10248–10253. doi: 10.1073/pnas.96.18.10248.

Varshney D, Spiegel J, Zyner K, Tannahill D, Balasubramanian S. 2020. The regulation and functions of DNA and RNA G-quadruplexes. Nat. Rev. Mol. Cell Biol. 21. doi: 10.1038/s41580-020-0236-x.

Villesen P, Mueller UG, Schultz TR, Adams RMM, Bouck AC. 2004. Evolution of ant-cultivar specialization and cultivar switching in Apterostigma fungus-growing ants. Evolution (N. Y). 58:2252–2265. doi: 10.1111/j.0014-3820.2004.tb01601.x.

Vogan AA et al. 2019. Combinations of Spok genes create multiple meiotic drivers in Podospora Rokas, A, Tautz, D, & Zanders, SE, editors. Elife. 8:e46454. doi: 10.7554/eLife.46454.

Wang H, Lambowitz AM. 1993. The mauriceville plasmid reverse transcriptase can initiate cDNA synthesis de novo and may be related to reverse transcriptase and DNA polymerase progenitor. Cell. 75:1071–1081. doi: https://doi.org/10.1016/0092-8674(93)90317-J.

Wang P et al. 2017. Frequent heteroplasmy and recombination in the mitochondrial genomes of the basidiomycete mushroom Thelephora ganbajun. Sci. Rep. 7:1–14. doi: 10.1038/s41598-017-01823-z.

Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. U. S. A. 84:9054–9058. doi: 10.1073/pnas.84.24.9054.

Wu B, Hao W. 2019. Mitochondrial-encoded endonucleases drive recombination of protein-coding genes in yeast. Environ. Microbiol. 21:4233–4240. doi: https://doi.org/10.1111/1462-2920.14783.

Wu C-S, Chaw S-M. 2015. Evolutionary Stasis in Cycad Plastomes and the First Case of Plastome GC-Biased Gene Conversion. Genome Biol. Evol. 7:2000–2009. doi: 10.1093/gbe/evv125.

Xu J, Yan Z, Guo H. 2009. Divergence, hybridization, and recombination in the mitochondrial genome of the human pathogenic yeast Cryptococcus gattii. Mol. Ecol. 18:2628–2642. doi: 10.1111/j.1365-294X.2009.04227.x.

Xu J, Zhang Y, Pun N. 2013. Mitochondrial recombination in natural populations of the button mushroom Agaricus bisporus. Fungal Genet. Biol. 55:92–97. doi: 10.1016/j.fgb.2012.09.004.

Yoshida W et al. 2018. Identification of G-quadruplex clusters by high-throughput sequencing of whole-genome amplified products with a G-quadruplex ligand. Sci. Rep. 8:3116. doi: 10.1038/s41598-018-21514-7.

Zhang D et al. 2020. Wolbachia limits pathogen infections through induction of host innate immune responses. PLoS One. 15:1–19. doi: 10.1371/journal.pone.0226736.

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. 209:1747–1756. doi: https://doi.org/10.1111/nph.13743.

Zhu Y et al. 2019. qDSB-Seq is a general method for genome-wide quantification of DNA double-strand breaks using sequencing. Nat. Commun. 10:2313. doi: 10.1038/s41467-019-10332-8.

Zug R, Hammerstein P. 2015. Bad guys turned nice? A critical assessment of Wolbachia mutualisms in arthropod hosts. Biol. Rev. 90:89–111. doi: https://doi.org/10.1111/brv.12098.

Several pictograms used in this thesis were obtained from the Noun Project website (https://thenounproject.com/, accessed 14-3-2021)

Summary

Mutualisms are abundant in nature despite our common understanding that evolution by natural selection is driven by competition. To understand how mutualisms can arise and what maintains their long-term stability, in this thesis I explore mutualisms at three levels of selection in the basidiomycete genus *Termitomyces*: the domestication of the fungus itself by a termite host, the inheritance and genetic variation of mitochondria in the fungal cell following sexual reproduction, and the transfer of tRNA genes from mitochondria to linear plasmids that results in co-dependency. In addition, I reconstruct the mitochondrial genomes of several *Termitomyces* and related species to analyze mtDNA variation and evolution in this diverse group of fungi.

The mitochondrial genomes of *Termitomyces* are characterized by a large, inverted repeat and a relatively high G4DNA motif content (Chapter 2). The significant enrichment of G4DNA in the repeats compared to the single copy regions suggests there may be reduced selection against G4DNA within the repeat regions, perhaps due to enhanced homologous recombination-driven repair of double-stranded breaks. G4DNA in fungal mtDNA is highly reduced in coding regions compared to human mtDNA, which may be the result of compensatory evolution in vertebrates to resolve G4DNA formation in transcribed RNA.

The common ancestor of *Termitomyces* was likely reliant on insect faeces as a growth substrate prior to its domestication by termites (Chapter 3). A phylogenomic reconstruction of *Termitomyces* and other Lyophyllaceae reveals that the free-living, insect-associated coprophiles *Arthromyces* and *Blastosporella* are close relatives of *Termitomyces*. Several key traits likely predate the fungus-termite mutualism, including conidiospores, a pseudorhiza, a perforatorium, and a repertoire of carbohydrate-active enzymes. The combination of these traits probably facilitated the transition of *Termitomyces'* free-living ancestor to a termite-domesticated lifestyle.

The mitochondria of several *Termitomyces* species seem to be dependent on linear mitochondrial plasmids for their supply of an essential tRNA gene (Chapter 4). Phylogenetic reconstruction of plasmids, mostly derived from *Termitomyces* species, reveals that transfer of mitochondrial tRNA genes from mtDNA to plasmids occurred independently in several *Termitomyces* species. In two cases, the tRNA gene was subsequently lost from the mtDNA of *Termitomyces*, the only known occurrences of complete loss of a tRNA function from *Termitomyces* mtDNA. This suggests an abrupt emergence of genetic addiction of mtDNA to a plasmid.

The mtDNA of *Termitomyces* shows limited recombination near regions containing mobile homing endonucleases (Chapter 5). Experimental matings of two *Termitomyces* homokaryon strains, along with complete mtDNA sequence data for six wild populations, suggest the inheritance of mtDNA proceeds without extensive heteroplasmy or genome-wide recombination. Previous studies noted *Termitomyces* appears to lack nuclear migration during mating, which would imply heterokaryon formation occurs in cells containing mixed cytoplasm. My results indicate that nuclear migration is either present but limited, or heteroplasmic cells quickly converge to a homoplasmic state. In either case, physical recombination of mtDNA is constrained but self-splicing endonucleases are able to transfer between different parental mitochondria during sex.

Acknowledgements

During this thesis project I have received help from many people for which I'm very grateful.

First of all, my daily supervisor Duur, thank you for entrusting me with this project and sharing your advice, enthusiasm and optimism for the past five years (and a bit). We both share a strong interest in not just evolutionary biology but also science education and philosophy, both of which you've helped me explore, for example during the Lorentz workshop and the Evoke project. Thank you for the freedom you've given me in this project and your unlimited support and patience.

Bas, thank you for all your support and input, especially during the regular project meetings. Those meetings were always a great source of help and motivation to me and your contributions were very valuable. I also want to thank you for your great work as chair of our group and always ensuring the atmosphere in the group is great.

Freek, when I decided to accept this thesis project, I knew I wanted you on board and I'm thankful you accepted. I built my experience in phylogenetics and genomics under your supervision during my Masters and you've continued to share your insights especially during the biweekly meetings. Thank you so much for all your help over the years.

Lennart, our PhDs have been co-dependent like the fungi and termites we study, and our theses are fused through a shared chapter. We've been able to help each other both in practical ways during lab and analytical work, but also with moral support when a venting session was required. In particular I fondly remember our trip to London, which is a great memory of many laughs and beards. I will never forgive you for teaching me the difference between single malt and blended scotch, as it costs a lot more now to stock my liquor cabinet. Nevertheless, it has been a privilege to work with you, and I am glad we are crossing the finishing line together.

Ben, you joined our group when my project was halfway, and you really brought a ton of knowledge and assistance to my and many other people's projects. Your keen eye, mind and voice have been invaluable to me and have improved my work considerably. Not only have you offered a bottomless well of knowledge and experience, but you were also always willing to help me out in the lab or with some bioinformatics problem. You are by far one of the most pleasant people I have ever worked with.

I want to thank all the other members of Team Duur: Eric, Sabine, Margo, Alex, and Anouk, for all the help I received and the great atmosphere in the group. The weekly meetings

were always fun and usually productive as well. I could also always count on your patient assistance when I was struggling in the lab on one of my experiments.

I want to thank the thesis students who participated in my project: Jeroen, Mart, Annetrude and Pjotr, for their contributions to some of the research in this paper, particularly Chapter 4. I also want to thank Fons for his advice on mitochondrial plasmids, which helped me out a lot when writing this Chapter.

I want to thank everyone in the Genetics department. It's an amazing community and I have missed it a lot working from home or an isolated office for the last year of my PhD. I won't name everyone here because many people have come and gone over the years, and I fear I might forget some names if I list them all. However, I will name some people I owe thanks to in particular. First of all, Wytske, thank you for sorting out lots of things for me and making my life a lot easier! I also appreciated you stopping by from time to time to ask how things were going. I also want to thank Frank, Corrie and José for helping me out in the lab with lots of things. Big thanks to Kim as well, for helping Lennart and me with the MinION sequencer, the great pub quizzes and many interesting discussions. Deep Space Nine is still better than Voyager though. I want to thank everyone on the GATC as well for hosting many legendary activities, and the vrijmibo gang for keeping things lively even during corona times with many ill-fated spacefaring adventures in Among Us. I also want to thank Arjan, Robin, Suzette and the other people involved with the first years Evolution & Systematics course, it was a lot of fun to assist with the practicals and the course was always a highlight of my year.

I want to give special thanks to my old friends from Biosystematics, where I first learned many of the skills I needed to complete this project. In particular, I want to thank Wilma, Lars, Sara, Floris, Eric and Tao for the occasional chat, advice, or bottle of home-brew beer.

I also want to give special thanks to Emma Salemink-Drenthel, for showing the way out of a dark forest.

Finally, I want to thank my friends and family, for all the love and support over the many years. Most of all, I want to thank my parents. Thank you for raising me to have an open but discerning mind, for encouraging my love for science and nature, and for always believing in me.

Cover design: Cover illustration by Marc F. P. M. Maas
Lay-out: Stefanie van den Herik | HerikMedia | www.herikmedia.nl
Printing: ProefschriftMaken | www.proefschriftmaken.nl

The research described in this thesis was financially supported by the Dutch Research Council (NWO VICI 86514007).

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.