Non-vertical genomics in fungal evolution

Miguel Ángel Naranjo Ortiz

Tesis doctoral UPF 2018

Director: Toni Gabaldón Estevan

Comparative Genomics Group
Bioinformatics and Genomics Department
Center for Genomic Regulation (CRG)

upf. Universitat Pompeu Fabra Barcelona

CRG Centre for Genomic Regulation

“Trust the fungus”

-Super Mario Brothers, the Movie

Acknowledgements

My relationship with the Fungi is a rather peculiar one, I would say. I remember playing Super Mario Bros. with my mother as a little kid and forcing myself to chomp sauteed mushrooms because I wanted to grow strong like the pixelated plumber. I want to thank my mom, who pushed me to keep studying in the face of adversity: You have supported me during my whole life and I cannot be thankful enough for this. You are just the strongest person I could ever imagine and this thesis would have never been had you not been there to tell me that everything will be fine the moment I needed the most. Thanks to my brother Alejandro too. I love you, bro. For the good old absurd laughs.

Like a good nerd kid growing in the 90's, I developed a mildly obsessive fixation with dinosaurs, that later switched toward arthropods. Knowing stuff became a major part of my identity as an individual, which in turn led me to decide that I wanted to study Biology. Fungi were there, for sure, but back then they were just a rather minor player. From that early times I want to thank my cousin Jose, and my good friends Álex and Michael.

Looking back, I think that my current obsession with mycology was just a matter of time. Botany was an amazing surprise during the first year on college, and microbiology was indeed an incredibly satisfying course. Even more, Edaphology became one of my favorite courses, with the exploration of such hyper complex system that is soil. These moments were shared with a select group of friends, with whom I have the fondest memories. Thanks to my university friends: Adrián, Ana, Sheila, Tamara and Mónica. Our paths diverged long ago, but I know we will keep meeting again and again.

Toward the end of my studies in the University of Alicante, a series of coincidences that include a handful of dried blueberries led me to the question: "Are Fungi monstrous enough to dedicate my whole life to them?". After a very brief pause, I decided that the answer was "Yes". That decision marked my entrance in the Fitolab, where I made a bunch of very good friends and I started my scientific productivity. Thanks to all the wonderful people that shared that experience with me. And to Nuria, who is the kindest and bravest person I know. I think that the only thing I can say is: I admire you.

I came to Barcelona to study a master and I shared my flat with some wonderful people. Thanks to Sara and Enric, who were there (late) to pick me up from the train station to my new home in a new land.. And of course, thanks to the "Tetra Master" members. To James, who became one of my best friends during a six hours long queue. To Fernando, unending source of crazy stories. And of course to Antonio, you lanky paranoid urban Mowgli. You know you are still on time to sell your soul to our fungal overlords.

Thanks to another unexpected coincidence I met Marina at a symposium, which in turn led me to enter the crazy bunch I have spent the last five years of my life with. Marina, you are at least half as good as a scientist as your are as a person. A quarter of any of those qualities would make anyone formidable. Thanks to the rest of the people that were with me in this ride. Thanks to Cinta and her never ending cheerfulness. Thanks to Ester for remember me so much when she's on holidays. To Pedro and Uciel, who shared the Colonies with me. To Jesse, who rescued me from the streets of Oeiras. To Rosa and Salva, who tried so hard to turn me into a real professional. To Laia, for singing with me "las pataticas". To Veronica, who helped me to plot the total annihilation of the human race. To Irene and to Ewa, who have shared so many coffees with me. To Hrant, who should work less often on weekends. To Ernst and his goat shenanigans. To them and to all the others I know I'm not mentioning, but for sure I haven't forgotten: Thank you very much.

And finally, to my boss and mentor Toni. I feel really fortunate to have been working for you and I am greatly honored to have been working with you. Thanks for everything.

Agradecimientos

Mi relación con los hongos es bastante peculiar, uno podría decir. Recuerdo jugar a Super Mario Bros. con mi madre de muy pequeño y forzarme a engullir champiñones salteados porque quería ser grande como el fontanero pixelado. Quiero agradecer a mi madre, que me empujó a seguir estudiando a pesar de las adversidades: Me has apoyado a lo largo de toda mi vida y no puedo esatr lo bastante agradecido por ello. Eres la persona más fuerte que puedo imaginar y esta tesis nunca habría sido posible si no hubieras estado allí para decirme que todo iba a salir bien cuando más lo necesitaba. Gracias también a mi hermano Alejandro. Te quiero, bro.

Como buen muchachito friki en los 90's, desarrollé una fijación tirando a obsesiva con los dinosaurios, que con el paso del tiempo viró hacia los artrópodos. Saber cosas se convirtió en una parte importante de mi propia identidad, lo que a su vez me llevó a decantarme por la Biología como carrera. De aquellos tiempos lejanos quiero agradecer a mi primo Jose y a mis buenos amigos Álex y Michael. Por las buenas risas absurdas.

Mirando atrás creo que mi actual obsesión por la micología era una cuestión de tiempo. La botánica fue una grata sorpresa durante mi primer año de universidad, y la microbiología resultó ser una asignatura increíblemente interesante. Después llegó la edafología, que se convirtió en una de mis asignaturas favoritas, con la exploración de un sistema tan hipercomplejo como es el suelo. Todo esto lo compartí con un selecto grupo de amigos, con los que tengo cientos de gratos recuerdos. Gracias a mis amigos de la universidad: Adrián, Ana, Sheila, Tamara y Mónica. Cada uno de nosotros tomó su camino hace ya tiempo, pero sé que seguiremos juntándonos una y otra vez.

Hacia el final de mis estudios en Alicante, una serie de coincidencias que incluyen un puñado de arándanos secos me llevaron a la pregunta: "¿Son los hongos lo bastante monstruosos como para dedicarles el resto de mi vida?". Tras una pausa muy breve decidí que la respuesta era "Sí". Y con esa decisión entré al Fitolab, donde hice otro buen puñado de amigos y comencé mi actividad como científico. Gracias a toda la gente que compartió conmigo aquella experiencia. Y gracias en particular a Nuria, que es la persona más amable y valiente que conozco. Creo que lo único que puedo decir es que te admiro.

Vine a Barcelona a estudiar el máster, y pasé a compartir piso con alguna gente maravillosa. Gracias a Sara y Enric, que estuvieron allí (tarde) para recogerme a la estación de tren y llevarme a esa nuevo hogar en esta nueva tierra. Y por supuesto, gracias a los miembros de "Tetra Master". A James, que se convirtió en uno de mis mejores amigos tras compartir una cola de seis horas. A Fernando, fuente inagotable de historias pintorescas. Y por supuesto a Antonio, Mowgli urbanita larguirucho y paranoico. Sabes bien que aún estás a tiempo de vender tu alma a nuestros señores fúngicos.

Gracias a otra casualidad conocí a Marina en un simposio, lo que a su vez me llevó a conocer a la jaula de grillos con la que he compartido estos últimos cinco años de mi vida. Marina, eres la mitad de buena como científica de lo que eres como persona. Y un cuarto de tan solo una de esas cualidad serían formidables en cualquiera. Y gracias a todos los demás que me han acompañado en este viaje. Gracias a Cinta y Susana su alegría sin fín. Gracias a Ester por acordarse tanto de mi cuando está de vacaciones. A Pedro y Uciel, que compartieron las Colonias conmigo. A Rosa y Salva, que pusieron mucho empeño que yo fuera un verdadero profesional. A Laia, por cantar conmigo "las pataticas". A Verónica, que me ayudó a planear la completa aniquilación de la raza humana. A Irene y Ewa, que han compartido tantos cafés conmigo. A Hrant, que debería trabajar menos los fines de semana. A Ernsts y sus locuras caprinas. A Álex, quien espero no se vengue por lo que hice en su defensa de tesis. A ellos y a todos los demás que sé que no he mencionado, pero seguro no he olvidado: Muchas gracias.

Y finalmente, gracias a mi jefe y mentor Toni. Ha sido una suerte poder trabajar para tí y un honor haber trabajado contigo. Gracias por todo.

Abstract

Non-vertical genomic events, which include horizontal gene transfer (HGT) and hybridization, are widely studied evolutionary processes in certain groups. HGT is particularly well known for prokaryotic organisms, while studies on hybridization started centuries ago for plants and animals. However, these phenomena are often overlooked when studying the evolution of other eukaryotic groups. Fungi, in particular, offer a privileged framework to study these kind of events, since the tractability of their genomes have made them target of hundreds of sequencing projects. This thesis reviews current knowledge on fungal diversity and evolutionary processes, with emphasis on the role of HGT and hybridization. We present an study on HGT of D-amino acid metabolic enzymes across eukaryotes, a novel computational framework for the study of hybrid genomes and other genomic anomalies, and a series of phylogenomic studies aimed to pinpoint possible ancient polyploidy events in fungi. Our findings expand current knowledge on non-vertical fungal evolution at different evolutionary and genomic scales, providing a multi-disciplinary perspective.

Resumen

Los eventos genómicos no verticales, que incluyen transferencia horizontal de genes (THG) e hibridación, han sido estudiados ampliamente en determinados grupos. THG es particularmente bien conocida para organismos prokariotas, mientras que los estudios sobre hibridación comenzaron en plantas y animales hace ya siglos. Sin embargo, estos fenómenos han sido a menudo ignorados a la hora de estudiar la evolución de otros grupos de eukariotas. Los hongos, en particular, ofrecen un sistema privilegiado en el que estudiar este tipo de eventos, puesto que la tractabilidad de sus genomas les ha llevado a ser objeto de cientos de proyectos de secuenciación. La presente tesis incluye una revisión del conocimiento actual acerca de la diversidad y procesos evolutivos, con énfasis en el papel de la THG y la hibridación. Presentamos un estudio acerca de THG de enzimas del metabolismo de D-amino ácidos en eukariotas, una nueva herramienta computacional para el estudio de genomas híbridos y otras anomalías genómicas, y una serie de estudios filogenómicos con la intención de identificar posibles eventos de paleopoliploidización en hongos. Nuestros resultados exapanden el conocimiento actual acerca de la evolución no vertical en hongos a diferentes escalas evolutivas y genómicas, proporcionando una perspectiva mulidisciplinar.

Contents

Preface

Miguel Ángel Naranjo-Ortiz, Toni Gabaldón

The present thesis deals with processes of non-vertical evolution in fungi, and how we can use genomic and phylogenomic methodologies to study them. While the number of examples of all of these phenomena keep stacking, current views state that these are collections of anecdotes, rather than proofs of widespread trends. Only a handful of studies have tried to infer global trends across fungi regarding their patterns of horizontal gene transfer (HGT), while none so far has tried to do the same for fungal hybrids and polyploidy. Because of this, my work has not been restricted to a specific fungal species, metabolic pathway, or specific biological process, but rather I have explored the growing diversity of fungal genomic sequences to identify particular cases that showcase the relevance of non-vertical evolution processes, hint at global evolutionary trends in the group, and showcase the applicability of the proposed methodologies. The evolutionary processes that I have focused on include horizontal gene transfer (HGT), hybridization, and polyploidization. All these processes have in common the breakage of traditional vertical inheritance rules, and thus can be detected based on the disruption they cause on gene phylogenies compared with the evolutionary history of the organism as a whole. HGT is the acquisition of genetic material from extraneous sources, a phenomenon that is accepted as ubiquitous in prokaryotes, but whose relevance in eukaryotic organisms is still under debate. Hybridization is the combination of two distant lineages by sexual or sexual-like means, producing a new lineage that presents mixed ancestry. Polyploidization is the presence of an abnormally high number of chromosomes in a given nucleus. Polyploidization is tightly linked to hybridization, as the later is one of the main causes of the former. Both hybridization and polyploidization have been thoroughly studied in plants and metazoans, but genetic peculiarities, technical limitations and historical disinterest have left the field severely lagging behind regarding the Fungi. For the detection and analysis of these evolutionary processes I have developed and used computational pipelines that comprise assembly and analysis of genomes and raw sequencing data, as well as large-scale phylogenetic analyses. Throughout my work I have extensively reviewed the literature of the field of fungal biology. This thesis includes a synthesis of the state of the art in the field, which focuses on the many impacts of comparative genomics and phylogenomics, together with some original theoretical ideas that we have developed in the process. Below I present a brief outline of the different chapters included in the thesis.

The Fungal kingdom is a highly diverse eukaryotic clade that includes mushrooms, toadstools, moulds, yeasts, lichens, rusts, smuts, and many other organisms that are less known by the general public (Richards *et al.* , 2017). Most fungi live as filamentous microbes, encased in a chitinous cell wall and feeding by absorbing the resulting metabolites (osmotrophy). Most fungi have lost their ancestral opisthokont flagellum, and virtually all of them are unable to perform phagocytosis. Yeasts are unicellular forms derived from filamentous growth and lichens form associations with photosynthetic Chlorophyta and Cyanobacteria. Early diverging Fungi are zoosporic and many live as parasites and parasitoids of planktonic algae. The kingdom is phylogenetically related to a

group of amoeboid protists (Nucleariida and Fonticulida), forming one of the two main branches within the Opisthokonta. The phylogeny and, consequentially, taxonomy of the clade has suffered radical changes that started with the first available gene sequences in the early 90s (Bowman *et al.* , 1992; Bruns *et al.* , 1992). Almost three decades later, the field has very efficiently exploited genomic technologies. Yet, numerous questions remain open, and the phylogenetic backbone of fungi is still not fixed. A robust phylogeny is essential for the analyses of HGT and comparative genomics **Chapter 1** of this thesis provides an up to date revision of the phylogeny and taxonomy of the kingdom Fungi.

The evolution of Fungi as a group is a greatly controversial topic, mostly due to the absence of a reliable fossil record. The oldest fungal fossils are approximately 400Mya, and represent either contemporary lineages or taxa whose affiliation cannot be easily assigned to any known clade (Berbee *et al.* , 2017). Compared with plants and animals, Fungi present several inherent cytological, genetic and evolutionary peculiarities that deserve their own dissertation. While fungal biology is an obscure subject for the general public, fungi are key components of virtually all ecosystems of our planet. Specially on land, Fungi act as decomposers, parasites, predators, symbiotic partners, and even primary producers, shaping ecosystems as we know them. As parasites of animals, Fungi are responsible of tremendous health burden in humans and cattle (Brown *et al.* , 2012). Disease causing fungi, in both plants and animals, have caused epidemic episodes that have inflicted severe losses to biodiversity (Fisher *et al.* , 2016). The role of Fungi in aquatic environments is less clear, although growing evidence suggests that fungi are much more diverse and play much more important roles in this ecosystem than previously thought (Richards *et al.* , 2012). **Chapter 2** of this thesis will review the main cytological and ecological transitions within fungi, with special emphasis on comparative genomics.

Fungi are an attractive clade for genomic studies in an era in which sequencing projects are affordable for even relatively small laboratories. As a consequence, the field of mycology has bloomed in the last three decades. However, these organisms have been shown to present a considerable genomic plasticity, including polyploidies, aneuploidies, chromosomal rearrangements, heterokaryosis and hybridization. The study of these processes is of great interest to understand the evolutionary mechanisms of the Fungi, but unfortunately add new layers of genomic complexity that are difficult to handle by non-specialized common computational approaches. The biochemical diversity of fungi is second only to that in bacteria, being able to exploit a wide array of substrates and produce a myriad of secondary metabolites. A good portion of this diversity owes to the acquisition of genes from other organisms via HGT (Marcet-Houben & Gabaldón, 2010; Richards *et al.* , 2011a; Wisecaver & Rokas, 2015). The mechanisms behind HGT are not fully understood and the relevance to the evolution of the group is still highly debated. In this regard, the very first limitation is the detection of the events. Phylogenetic approaches can be plagued by tree reconstruction artifacts, incomplete lineage sorting, gene gain and loses, defective gene annotation and architectural changes (Dupont & Cox, 2017; Wisecaver & Hackett, 2014). Even more, putative events must prove they are genes contained within the real genome of the organism, rather than contaminating sequences. **Chapter 3** of this thesis reviews the evolutionary and methodological implications of these phenomena in Fungi. **Chapter 4** will present a phylogenetic study on HGT of enzymes related to D-amino acid metabolism across all eukaryotes.

Chapter 5 presents Karyon, a novel computational tool developed to diagnose problematic *de novo* assemblies, an approach we believe will be useful for many genomic studies in microbial eukaryotes. **Chapter 6** will present the results of our phylogenomic analyses within the context of 1000 Fungal Genomes Project, in which we have used phylogenomic approaches to solve the phylogenetic relationship among groups within Taphrinomycotina and the phylogenetic placement of Wallemiomycetes within Basidiomycota. We have also analyzed patterns of gene duplication across the fungal tree of life in order to pinpoint putative recent and ancestral polyploidization events. Finally, **Chapter 7** presents the genome sequence and analyses of *Monilinia laxa*, a fungal pathogen of diverse species of orchard trees in the Rosaceae.

References

Berbee, Mary L., James, Timothy Y., & Strullu-Derrien, Christine. 2017. Early Diverging Fungi: Diversity and Impact at the Dawn of Terrestrial Life. *Annual Review of Microbiology*, **71**(1), 41–60.

Bowman, B H, Taylor, J W, Brownlee, A G, Lee, J, Lu, S D, & White, T J. 1992. Molecular evolution of the fungi: relationship of the Basidiomycetes, Ascomycetes, and Chytridiomycetes. *Molecular biology and evolution*, **9**(2), 285–96.

Brown, Gordon D, Denning, David W, Gow, Neil A R, Levitz, Stuart M, Netea, Mihai G, & White, Theodore C. 2012. Hidden killers: human fungal infections. *Science translational medicine*, **4**(165), 165rv13.

Bruns, T D, Vilgalys, R, Barns, S M, Gonzalez, D, Hibbett, D S, Lane, D J, Simon, L, Stickel, S, Szaro, T M, & Weisburg, W G. 1992. Evolutionary relationships within the fungi: analyses of nuclear small subunit rRNA sequences. *Molecular phylogenetics and evolution*, **1**(3), 231–41.

Dupont, Pierre-Yves, & Cox, Murray P. 2017. Genomic Data Quality Impacts Automated Detection of Lateral Gene Transfer in Fungi. *G3: Genes, Genomes, Genetics*, **7**(4).

Fisher, Matthew C, Gow, Neil A R, & Gurr, Sarah J. 2016. Tackling emerging fungal threats to animal health, food security and ecosystem resilience. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **371**(1709).

Marcet-Houben, Marina, & Gabaldón, Toni. 2010. Acquisition of prokaryotic genes by fungal genomes. *Trends in genetics : TIG*, **26**(1), 5–8.

Richards, Thomas A., Leonard, Guy, Soanes, Darren M., & Talbot, Nicholas J. 2011a. Gene transfer into the fungi. *Fungal Biology Reviews*, **25**(2), 98–110.

Richards, Thomas A., Jones, Meredith D.M., Leonard, Guy, & Bass, David. 2012. Marine Fungi: Their Ecology and Molecular Diversity. *Annual Review of Marine Science*, **4**(1), 495–522.

Richards, Thomas A., Leonard, Guy, & Wideman, Jeremy G. 2017. What Defines the "Kingdom" Fungi? *Microbiology Spectrum*, **5**(3).

Wisecaver, Jennifer H, & Hackett, Jeremiah D. 2014. The impact of automated filtering of BLAST-determined homologs in the phylogenetic detection of horizontal gene transfer from a transcriptome assembly. *Molecular Phylogenetics and Evolution*, **71**, 184–192.

Wisecaver, Jennifer H., & Rokas, Antonis. 2015. Fungal metabolic gene clusters: caravans traveling across genomes and environments. *Frontiers in Microbiology*, **6**(mar), 161.

Chapter 1

Diversity, taxonomy and phylogeny of the Fungi

Miguel Ángel Naranjo-Ortiz, Toni Gabaldón

Abstract

The fungal kingdom comprises a hyperdiverse clade of heterotrophic eukaryotes characterized by the presence of a chitinous cell wall, the loss of phagotrophic capabilities and cell organizations that range from completely unicellular monopolar organisms to highly complex syncitial filaments that may form macroscopic structures. Fungi emerged as a "Third Kingdom", embracing organisms that were outside the classical dichotomy of animals versus vegetals. The taxonomy of this group has a turbulent history that is only now starting to be settled with the advent of genomics and phylogenomics. We here review the current status of the phylogeny and taxonomy of fungi, providing an overview of the main defined groups. Based on current knowledge, nine phylum-level clades can be defined : Opisthosporidia, Chytridiomycota, Neocallimastigomycota, Blastocladiomycota, Zoopagomycota, Mucoromycota, Glomeromycota, Basidiomycota and Ascomycota. For each group, we discuss their main traits and their diversity, focusing on the evolutionary relationships among the main fungal clades. We also explore the diversity and phylogeny of several groups of uncertain affinities and the main phylogenetic and taxonomical controversies and hypotheses in the field.

Naranjo-Ortiz MA, Gabaldón T. Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. Biol Rev. 2019 Dec 29;94(6):2101–37. DOI: 10.1111/brv.12550

Chapter 2

Ecology and evolution of Fungi

Miguel A. Naranjo-Ortiz, Toni Gabaldón

Abstract

Fungi are a highly diverse group of heterotrophic eukaryotes characterized by the absence of phagotrophy and presence of chitinous cell wall. While unicelular fungi are far from rare, part of the evolutionary success of the group resides in their ability to grow indefinitely as a cylindrical multinucleated cell (hypha). Armed with these morphological traits and with a extremely high metabolical diversity, fungi have conquered numerous ecological niches and have shaped a whole world of interactions with other living beings. In this review we will survey the main evolutionary and ecological processes that have guided fungal diversity. In this regard, we will first review the ecology and evolution of the zoosporic lineages and the process of terrestrialization, as one of the major evolutionary transitions in this kingdom. Several plausible scenarios have been proposed for fungal terrestralization and we here propose a new one, which considers icy environments as a transitory niche between water and emerged lands. We will then focus on exploring the main ecological relationships of Fungi with other organisms (Other fungi, protozoans, animals and plants), as well as the origin of adaptations to certain specialized ecological niches within the group (lichens, black fungi and yeasts). Throughout the review we will use an evolutionary and comparative genomics perspective to understand fungal ecological diversity.

Naranjo-Ortiz MA, Gabaldón T. Fungal evolution: major ecological adaptations and evolutionary transitions. Biol Rev. 2019 Apr 25;94(4):brv.12510. DOI: 10.1111/brv.12510

Chapter 3

Non vertical genomics in Fungi

Miguel A. Naranjo-Ortiz, Toni Gabaldón

Abstract

The present chapter will review the concepts of non-vertical genetics and how they affect fungi. Non-vertical or reticulated evolution includes instances in which genetic material of a given lineage merges with another lineage, and includes mainly the phenomena of horizontal gene transfer (HGT) and hybridization. Additionally, in order to properly discuss these phenomena we must discuss current knowledge on genome architecture evolution in fungi, which includes chromosomal rearrangements and heterokaryosis.

Horizontal gene transfer (HGT) is a common phenomenon in fungi, related to many evolutionary transitions. The exact scope of this form of gene flow is poorly explored, though. Most studies focus on HGT across large evolutionary distances, which provide the most spectacular examples. However, transference of genes among closely related strains must be highly common. Detection of such events is plagued by methodological limitations, ranging from the lack of phylogenetic signal, genome assembly artifacts of transposon-rich regions that tend to contain these genes and limitations in genome annotation algorithms. This implies that fungi must be understood under a pangenome-based approach that has been so prevalent for prokaryotes in the last decades. The study of fungal pangenomes is in its earliest infancy, but the topic is experimenting a meteoric rise. Here we review the current state of horizontal gene transfer in Fungi as well as the recent bibliography regarding fungal pangenomes

Compared to other eukaryotes, the Fungi present a very compact and plastic genomic organization that grants them a huge adaptability. In this section we review how fungi use drastic chromosomic rearrangements, mainly aneuploidies and polyploidies, to accelerate the adaptation to novel environmental conditions and the evolutionary trade-offs they must face when doing so. Like in plants, one important source of genome instability in fungi is the emergence of hybrid lineages. However, our knowledge in the subject lags one century behind of plants, having only been able to start study it with the advent of sequencing technologies. We discuss known examples of fungal hybrids, the peculiarities of hybridization in this group and the methodological approaches to the study of fungal hybridization. Finally, many fungi live as syncytial entities. The existence of multiple non-identical nuclei within a single cytoplasm adds another layer of genome plasticity and opens up the possibility of intra-cytoplasmic ecological interactions. Intra-cytoplasmic ecology is a classic problem in biology that we consider was forgotten by the community before the emergence of the tools to properly study it. Thus, we will try to contextualize heterokaryotic conflict under the light of the genomic era.

3.1 New tricks for old dogs: Horizontal gene transfer in Fungi

After the publication of the first human genome drafts, the field of horizontal gene transfer (HGT) in eukaryotes entered its "Dark Age". The human genome paper included the claim of more than two hundred genes putatively acquired from bacteria via HGT (Lander *et al.* , 2001). This finding was later discarded as being an artifact resulting from insufficient sampling (Salzberg, 2017; Stanhope *et al.* , 2001). This led to a decade in which HGT detection in eukaryotic genomes was faced with severe skepticism, with most genome studies not even exploring that possibility. Yet, evidence started to accumulate that HGT was not an uncommon event in some eukaryotic lineages, including fungi. The detection of HGT events usually requires finding incongruences between a gene phylogeny and the known species tree (Galtier & Daubin, 2008; Grant & Katz, 2014; Katz, 2015; Leigh *et al.* , 2011; Naranjo-Ortíz *et al.* , 2016; Nguyen *et al.* , 2015b; Soucy *et al.* , 2015; Szöllősi *et al.* , 2015). This requirement has been traditionally difficult to meet for most eukaryotic lineages, since genomic studies have long been focused on just a few economically important species, limiting the taxonomic coverage. However, fungi were an early exception to this. With a taxonomically robust backbone of sequenced genomes and DNA coming from axenic cultures, HGT claims in Fungi are usually quite robust. Once it was possible to search for HGT in a wide taxonomical range of Fungi, it was revealed that this evolutionary phenomena affects Pezizomycotina preferentially (Gluck-Thaler *et al.* , 2015; Marcet-Houben & Gabaldón, 2010; Wisecaver *et al.* , 2014), at least when compared with Saccharomycotina and Basidiomycota. Furthermore, in some cases HGT have been shown to have been key in the acquisitions of important adaptations, as shown in Neocallimastigomycotina and its outstanding carbohydrate degrading enzimatic pool (Garcia-Vallvé *et al.* , 2000; Rosewich & Kistler, 2000).

Despite the increasing anecdotal evidence for HGT in Fungi, a capital question remains unsolved. The exact mechanism that allows transference of genetic material into the genome of a fungus is not understood (Andersson, 2005,0; Richards *et al.* , 2006,1; Soanes & Richards, 2014). Since fungal cells are surrounded by a thick cell wall and do not have phagotrophic capabilities, ingested microbes cannot be a source of HGT, as posited by the "you are what you eat" hypothesis(**?**). Reports of in vitro recombination between bacteria and yeasts have been published (Heinemann & Sprague, 1989,9; Inomata *et al.* , 1994; Moriguchi *et al.* , 2013; Sawasaki *et al.* , 1996; Suzuki *et al.* , 2015). On the other hand, *Agrobacterium*-like bacteria can be used to transform filamentous fungi in laboratory, suggesting these soil microbes might be to blame (Jiang *et al.* , 2013; Lacroix & Citovsky, 2016; Lacroix *et al.* , 2006). While the mechanisms are not fully understood, the content of HGT events is better known. Horizontally transmitted genes obey the "complexity hypothesis" (Jain *et al.* , 1999), that describes the gene content of an organism in terms of network connectivity. A new element added to such network must be able to function in a relatively isolated way. In practice, this implies that simple transporters, simple enzymatic pathways or genes in the secondary metabolism are more likely to be transferred. Some types of enzymes, such as amino acid racemases seem particularly prone to HGT in fungi and other microbial eukaryotes, although their physiological role in the receiving organisms remains to be clarified Marcet-Houben & Gabaldón (2010); Naranjo-Ortíz *et al.* (2016). At the other extreme, genes related to information processing, such as translation or transcription, are among the most recalcitrant genes for HGT. Comparative genomic studies support the paradigm of complexity hypothesis Marcet-Houben & Gabaldón (2010); Wisecaver & Rokas (2015); Wisecaver *et al.* (2014).

There are good arguments suggesting that a fair portion of the HGT events are simply not detected. In general, it is far easier to detect an event if the resulting phylogenetic incongruity is large, such as that caused by inter-domain transfers Galtier & Daubin (2008); Haegeman *et al.* (2014); Leigh *et al.* (2011); Marcet-Houben & Gabaldón (2010); Naranjo-Ortíz *et al.* (2016). However, HGT among fungi has been proved. In the hypothetical scenario of HGT between relatively close species (for instance, members of the same family), alternative hypotheses of differential gene loses or incomplete lineage sorting cannot be ruled out. Ancient HGT events might have an untraceable phylogenetic signal, something to be expected considering these genes must necessarily be under natural selection or that databases might lack adequate representation of the donor group. HGT

might occur several times independently for the same gene, either through sequential transfers or by independent transfers from a phylogenetically close donor. This, in turn, would produce complex phylogenetic patterns that can prove difficult to interpret. On top of that, certain protein families evolve in ways that make molecular phylogenies intricate and difficult to resolve. For instance, non-ribosomal peptide synthases (NRPS) are modular multidomain enzymes that can evolve by losing, duplicating or shuffling functional domains, rather than by point mutations (Hur *et al.* , 2012; Marahiel, 2009; Weber & Marahiel, 2001). Many protein families display extremely low sequence conservation, which difficult the reconstruction of accurate molecular phylogenies, even for small evolutionary distances (Ponting, 2017). This is true, for instance, for secreted peptidases in Fungi (Krishnan *et al.* , 2018; Poppe *et al.* , 2015). Both NRPS and peptidases are well known virulence factors in many fungi or have relevant roles for certain niches, are compatible with the complexity hypothesis, and have a huge diversity in filamentous fungi, making them prime candidates for HGT events.

HGT can play an important role in adaptation of species to novel niches or lifestyles, and has been proven to be involved in short term evolutionary transitions. For instance, the acquisition of a toxin-encoding gene by *Pyrenophora tritici-repentis* from *Stagonospora nodorum* turned a fungus causing occasional spots in wheat leaves into a devastating pest in a matter of decades (Friesen *et al.* , 2006; Oliver & Solomon, 2008). HGT between plant pathogens seems to be fairly widespread and, in certain cases, patterns of repeated HGT between the same groups have been reported (Armijos-Jaramillo *et al.* , 2014; Bettini *et al.* , 2014; Gluck-Thaler *et al.* , 2015; Khaldi *et al.* , 2008; Qiu *et al.* , 2016; Yin *et al.* , 2016). This mounting evidence suggests that HGT is an important factor promoting the rise of novel plant pathogens. Another example of recent acquisition of novel genes with functional implications implies fungi involved in beverage and food production environments. Wine strains of *Saccharomyces cerevisiae* contain genes that are not present in beer strains and are probably related to adaptation to their particular industrial environment (Borneman *et al.* , 2011; Galeote *et al.* , 2010; Novo *et al.* , 2009). Analogously, several *Penicillium* species growing on cheese have been shown to contain recently acquired genes that are adaptive to such particular environment (Ropars & Corradi, 2015). These examples highlight the power of HGT in enabling microbial adaptation to novel niches, including those involved in domestication.

While ancient events are difficult to detect, there is clear evidence that some fungal groups have been affected by HGT over long evolutionary scales. Many yeasts in the Saccharomycotina present an horizontally transferred cytoplasmic dihydroorotate dehydrogenase that allows for the biosynthesis of pyrimidines in anoxic conditions (Gojkovic *et al.* , 2004). HGT of high affinity nitrate transporters from Oomycota has been proposed to be an important feature for the land hegemony of Dikarya (Slot & Hibbett, 2007). A large fraction of the polyketide synthase genes in Lecanoromycetes seem to arise from Actinobacteria, followed by gene expansions (Schmitt & Lumbsch, 2009). Some of these genes are involved in the synthesis of ecologically relevant secondary metabolites such as mycotoxins or antibiotics. However, some studies suggest that these are exceptional findings, and that, overall, HGT events in eukaryotes are evolutionarily short-lived (Katz, 2015).

3.2 Fungal pangenomes

The concept of pangenome emerges from the realization that the gene content differs among strains or isolates within a defined taxonomic unit, usually a species (Tetz, 2005). Similar to HGT, the concept has been widely studied in Bacteria. The pangenome refers to the complete gene pool of a single species, including copy number variations, horizontally acquired genes and mobile genetic elements, such as plasmids. For instance the different strains of *Escherichia coli* have genomes typically containing 3500-4000 protein coding genes, but with more than 2000 sequenced strains, the known pangenome of the species contains over 18000 genes (Jang *et al.* , 2017). The concept can be applied to eukaryotes as well, including plants (Golicz *et al.* , 2016a,1; Springer *et al.* , 2009) and fungi. The pangenome can be divided into two distinct subsets. The core genome refers to the set of genes that are universally present in all analyzed genomes of a given species, while the variable (also known as accessory or flexible) genome comprises the rest of genes which can be

absent or present, depending on the analyzed strain. Obviously, the composition of the pangenome and the core and accessory subsets depend on the number of analyzed strains

Pangenome studies require the sequence and comparison of multiple strains of the same species, and no other fungus has been studied in more detail than *Saccharomyces cerevisiae* (Engel & Cherry, 2013; Gallone *et al.* , 2016; Strope *et al.* , 2015a). Compared with other Fungi, specially with moulds, the secretome and secondary metabolism of yeast is very limited. Even then, a significant number of genes are not fixed in the global yeast population. Comparative analyses of 100 strains characterized the patterns of presence/absence variation of several genes associated to highly variable phenotypic traits, such as resistance to copper, sulfite or cycloheximide (Strope *et al.* , 2015a). Some of these genes seem to have been acquired by introgression from other *Saccharomyces* species, mainly *S. paradoxus*; and quite often present copy number variation across isolates. A recent study over more than 1000 *S. cerevisiae* strains coming from diverse environments have provided a reliable pangenome estimation of almost 7800 genes; from which approximately 4940 are core and 2860 are accessory (Peter *et al.* , 2018). Accessory genes tend to concentrate in subtelomeric regions, and are enriched in cell-cell interactions, secondary metabolism and stress responses. Also, these genes have a higher tendency to appear in copy number variation and hemizygosity. Wine yeasts present high tendency to form hybrids, and some HGT associated to adaptation to this niche have been described (Borneman *et al.* , 2011; Galeote *et al.* , 2010; Marsit *et al.* , 2015). The same can be said about the sake strain K7. This strain is highly similar to the reference S288c except for the presence of certain subtelomeric regions that include up to 49 ORFs that are absent in the reference, as well as 49 genes that are absent in K7 (Akao *et al.* , 2011). On the other hand, sequencing of the laboratory strain CEN.PK113-7D revealed absence of 83 genes that are present in the S288c, and a small number of genes that appeared only in CEN.PK113-7D (Nijkamp *et al.* , 2012). Among these last batch there was the surprising finding of a functional biotin biosynthetic pathway (Nijkamp *et al.* , 2012). All these studies showcase that *S. cerevisiae* has a relatively small pangenome that, nonetheless, has a very profound effect in shaping the phenotypic diversity of the species. It should be noted that most of these studies treat yeast hybridization as an anomaly, when we should consider the scope of yeast hybrid diversity as part of the total genetic pool of the species. *S. cerevisiae* has some peculiarities compared to other yeast species, specially regarding the frequency of sexual mating (Dujon & Louis, 2017; Hittinger, 2013; Ni *et al.* , 2011; Zeyl, 2009), which probably makes it a poor model for the whole Saccharomycotina in regards of pangenome dynamics.

Pangenomes have been studied for a handful of filamentous Ascomycota. *Beauveria bassiana* is an entomopathogenic member of the Hypocreales (Sordariomycetes) that presents a wide host range. Valero-Jiménez et al, compared five isolates with varying virulence phenotypes against mosquitoes (Valero-Jiménez *et al.* , 2016). For *B. bassiana*, the core genome contained around 7300 genes, while the pangenome size was estimated at around 13000 genes. In the case of the most virulent strain, 163 genes were strain specific, mostly containing secondary metabolic clusters located near telomeric regions. It is noteworthy that some of these strain exclusive genes have homology to other entomopathogenic Hypocreales or bacteria, suggesting an origin by horizontal gene transfer. *Aspergillus* and *Penicillium* are two widely studied genera of Eurotiales (Eurotiomycetes). These genera produce a great diversity of secondary metabolites, which can be traced to the wide range of genome and proteome sizes across both genera (de Vries *et al.* , 2017; Nielsen *et al.* , 2017). There are already several comparative genomic studies that focus on strains with different biosynthetic capabilities, although most of them use re-sequencing approaches that greatly limit the ability to find strain-specific genes. The few that take specific methods to address this question show that each strain carries a significant number of specific genes (Gilbert *et al.* , 2018; Ropars *et al.* , 2015). Pangenomic studies are of great relevance to understand the emergence of virulent traits in plant pathogens. In this regard, accessory chromosomes in *Fusarium spp.* and *Zymoseptoria tritici*, are a well known form of pangenome. *Fusarium* in particular has been shown to present a mitochondrial pangenome (Brankovics *et al.* , 2018), with evidence of mitochondrial recombination between strains. Comparison of subtelomeric regions among 6 strains of *Fusarium fujikuroi* reveal the presence of lineage-specific secondary metabolic pathways, some of which have emerged from HGT events (Chiara *et al.* , 2015). Analysis of 60 *F. graminearum* isolates form North America identified a pangenome close to 1700 genes that seem to concentrate in AT-rich regions, probably

centromeres or telomeres (Kelly & Ward, 2018). Similar results have been obtained for *F. fujikuroi* (Chiara *et al.* , 2015; Niehaus *et al.* , 2017), and *F. oxysporum* (Armitage *et al.* , 2017a). All in all, these studies have focused either on geographically delimited isolates or in comparison of a handful of isolates. The pangenome of the wheat pathogen *Z. tritici* has recently been extensively studied (Plissonneau *et al.* , 2016,1) and it is probably the only filamentous fungus species for which a reasonably complete pangenome has been described so far. *Z. tritici* contains a pangenome comprising more than 17,400 protein coding genes. All analyzed isolates possessed around 12,000 genes, while the core genome for all analyzed strains lies around 9,100 genes. This fungus presents large effective population sizes and frequent genetic exchange between populations (Zhan *et al.* , 2003). As mentioned in the case of *Aspergillus* and *Penicillium*, the development of a proper pangenomic paradigm on fungal genomics have been severely limited by the use of re sequencing technologies in eukaryotes, where a single reference genome is used to map sequencing libraries coming from the rest. Applying *de novo* sequencing techniques to all strains in a given study is much more expensive, but should still be realistic for many laboratories. However, this approach can be very limited by the fact that accessory genes tend to concentrate in AT-rich regions such as telomeres, centromeres, genomic islands and transposable element rich regions, which in turn tend to be very difficult to assemble. As an alternative, RNAseq might be used to identify low coverage genes that are probably absent in the strain, as well as novel transcript without homology in the reference. These genes tend to evolve at a faster rate than the rest of the genome (Dong *et al.* , 2015), which of course further limits homology based annotation approaches. In any case, as more fungal genomes accumulate, the concept of pangenome is starting to become more widely adopted, and the field is going to experiment a substantial growth in the following years. The exposed examples show that pangenomes greatly vary in size and functional relevance across different fungal lineages. The existence of an extensive pangenome requires efficient mechanisms for genetic exchange, which may or may not be of sexual nature. Our current knowledge in this area is severely lacking. Sex is only indirectly inferred from genomic data for many groups and mechanisms of HGT in Fungi are still unknown. Regarding the later, pangenomes forces to reevaluate the hypothesis that HGT drives gene clustering in Fungi, as HGT among closely related Fungi might be much more prevalent than previously thought. In this regard, it is important to point the possible relation between mesosynteny and pangenomes (Hane *et al.* , 2011). Mesosynteny seems to be particularly prominent among Dothideomycetes (Hane *et al.* , 2011), the group to which *Z. tritici* belongs.

The pan-genome of the model arbuscular mycorrhizic fungi *Rhizophagus irregularis* has been studied only recently, but the results of these works are revolutionary (Mathieu *et al.* , 2018). After comparing just six laboratory strains, the results are that only 50% of the genes are shared among all of them, and the remaining, accessory fraction contains over 150,000 genes (Chen *et al.* , 2018a). It remains to be seen how these results can be interpreted in the context of the heterokaryotic nature of these fungi. First of all, the existence of such huge pan-genome can only be explained in the light of rampant gene flow, either by sexual or sexual-like interchange among strains or, alternatively, by large ratios of HGT. Second, if these fungi are heterozygous, what are the implications for their intracytoplasmic pangenomes? How are these accessory genes distributed across the nuclear population within a certain mycelium? If distinct nuclei present accessory genes that are not present in all the cytoplasmic population, this would greatly affect genome assembly and annotation, as the sequencing depth would be highly variable in these regions and reads spanning these regions would be incongruent. The size of this pan-genome might as well be in the root of the fact that arbuscular mycorrhizic fungi present an species diversity and endemicity that is orders of magnitude lower than that of land plants, their obligate hosts. *R. irregularis* is basically the only member of Glomeromycota that can be cultured in laboratory. Assuming that the findings in this species can be extrapolated to the rest of the phylum, the genetic pool accessible to these fungi may be truly enormous (Mathieu *et al.* , 2018).

3.3 Chromosome poker: Aneuploidy and polyploidy in Fungi

Plants and animal require to finely tune their gene dosage and the relative expression level of the different components of their signaling pathways (Oromendia *et al.* , 2012; Otto, 2007; Schoenfelder

& Fox, 2015; Soltis *et al.* , 2015). Deviations from the optimal ploidy conditions are commonly associated with grave inborn defects. Plants and some animals seem to, at least tolerate polyploidy, since it keeps the stoichiometry of the gene products; although it might still cause reproductive defects. However, many exceptions exist to this rule when looking at particular cell types, such as *Drosophila* salivary glands, human hepatocytes or extraembryonary tissues in plants; all of which present aneuploidies or polyploidies under physiological conditions (Otto, 2007; Schoenfelder & Fox, 2015; Torres *et al.* , 2008). Tumoral cells, which are undoubtedly "fitter" than their healthy counterparts, tend to present numerous chromosome aberrations. Thus, the grave phenotypical defects caused by chromosomal aberrations seems not to be such for cells themselves, and the paradigm should rather be interpreted in the context of organogenesis. Freed from this preconceived ideas, we must face the next obvious question. What is the case in organisms that, just like fungi, live without organs?

Aneuploidy is the presence of different ploidy levels within the same genome, usually affecting entire chromosomes or large chromosomic regions. Aneuploidies emerge spontaneously within populations (Torres *et al.* , 2008) and tend to have drastic effects on the fitness due to imbalance in gene dosage and formation of multipolar meiotic and mitotic divisions (Bonney *et al.* , 2015; Dodgson *et al.* , 2016; Kumaran *et al.* , 2013; Oromendia *et al.* , 2012; Torres *et al.* , 2008). However, under certain conditions aneuploidies might provide selective advantages (Bennett *et al.* , 2014; Berman, 2016; Gerstein *et al.* , 2015; Kravets *et al.* , 2014; Todd *et al.* , 2017). For instance, if a fungus is trying to grow in a medium containing a toxic compound or a rare nutrient source, aneuploid cells that increase gene dosage for genes related to detoxification of such toxin or exploitation of such nutrient would probably tolerate higher concentrations and would be fitter than their euploid counterparts. This dynamic is observed in fungal pathogens acquiring resistance to antifungal drugs (Anderson *et al.* , 2017; Farrer *et al.* , 2013; Harrison *et al.* , 2014; Morrow & Fraser, 2013; Sionov *et al.* , 2010; Sun *et al.* , 2014). While the same phenotypic effect would be possible by tandem duplications of the genes, the frequency of these mutations is much lower. Aneuploidies can easily revert to an euploid state if the stressful condition is transitory. If no reversion occurs, additional mutations that reduce the negative effects of the altered ploidy state while keeping the selected phenotype will be selected. Thus, aneuploidies serve as transitory intermediate states during the process of adaptation to novel conditions (Anderson *et al.* , 2017; Berman, 2016; Berman *et al.* , 2016; Farrer *et al.* , 2013; Harrison *et al.* , 2014; Morrow & Fraser, 2013). Finally, aneuploidies themselves might serve as compensatory mutations for polyploidies. In this regard, the best studied example is the parasexual cycle of *Candida albicans* (Saccharomycotina) (Brown *et al.* , 2014; Forche, 2012; Forche *et al.* , 2008; Hickman *et al.* , 2015; Whelan *et al.* , 1985). This process is a bit different to the mating in filamentous fungi that lead to the formation of a heterokaryon, also known as parasexual cycle (Daskalov *et al.* , 2017; Pontecorvo, 1956). In *C. albicans* fusion of two diploid cells by non meiotic mating results in an effective tetraploid state. The tetraploid is genomically unstable, and suffers concerted chromosome loss that recovers stability. Thus, parasexual cycle promotes aneuploidies and helps this pathogenic yeast to adapt to the host immune system, as well as pharmacological treatments (Bennett *et al.* , 2014; Gerstein & Berman, 2015; Harrison *et al.* , 2014). Parasexual cycle in *C. albicans* has apparently evolved through loss of part of the meiotic recombinatory machinery, but aneuploidy might still be an important genome stabilizer in instances in which meiotic recombination is impaired, such as hybridization events (see below). Aneuploid populations are well known also for other Saccharomycotina isolated from industrial environments, such as *Saccharomyces* (Borneman *et al.* , 2011; Walther *et al.* , 2014; Zhu *et al.* , 2016) or *Brettanomyces* (Avramova *et al.* , 2018; Borneman *et al.* , 2014; Hellborg & Piskur, 2009); as well as the frog killing chytrid *Batrachochytrium dendrobatidis* (Joneson *et al.* , 2011; Rosenblum *et al.* , 2013a).

Just like plants, fungi are able to undergo autopolyploidization, in which all the chromosomes have the same genotype at the moment of the duplication; or allopolyploidization, in which the chromosomes are genetically distinct (Albertin & Marullo, 2012; Otto, 2007; Todd *et al.* , 2017). At mid evolutionary term, these events provide plenty opportunities for innovation through extensive gene duplication followed by subfunctionalization and neofunctionalization (Albertin & Marullo, 2012; Conant & Wolfe, 2008; Magadum *et al.* , 2013). However, autopolyploidization does not provide genotypic innovation at the short term, a necessity for fixing this mutation within a population.

Despite the lack of genotypic novelty, autopolyploidy might still provide some advantages. This process usually reduces surface-volume ratio, producing larger cells (Otto, 2007; Schoenfelder & Fox, 2015). This change has great implications for membrane transport, which in turn affects general metabolism. An increased size might be selectively advantageous against certain pressures, such as phagocytic predation. This seems to be the case for titan cells in *Cryptococcus*, which are polyploid vegetative cells resistant to the attack of the immune system and from which infectious diploid and aneuploid cells emerge (Gerstein *et al.* , 2015; Okagaki & Nielsen, 2012). Finally, a putative tetraploid state has been described for a widespread strain of the microsporidian *Nosema ceranae* (Pelin *et al.* , 2015).

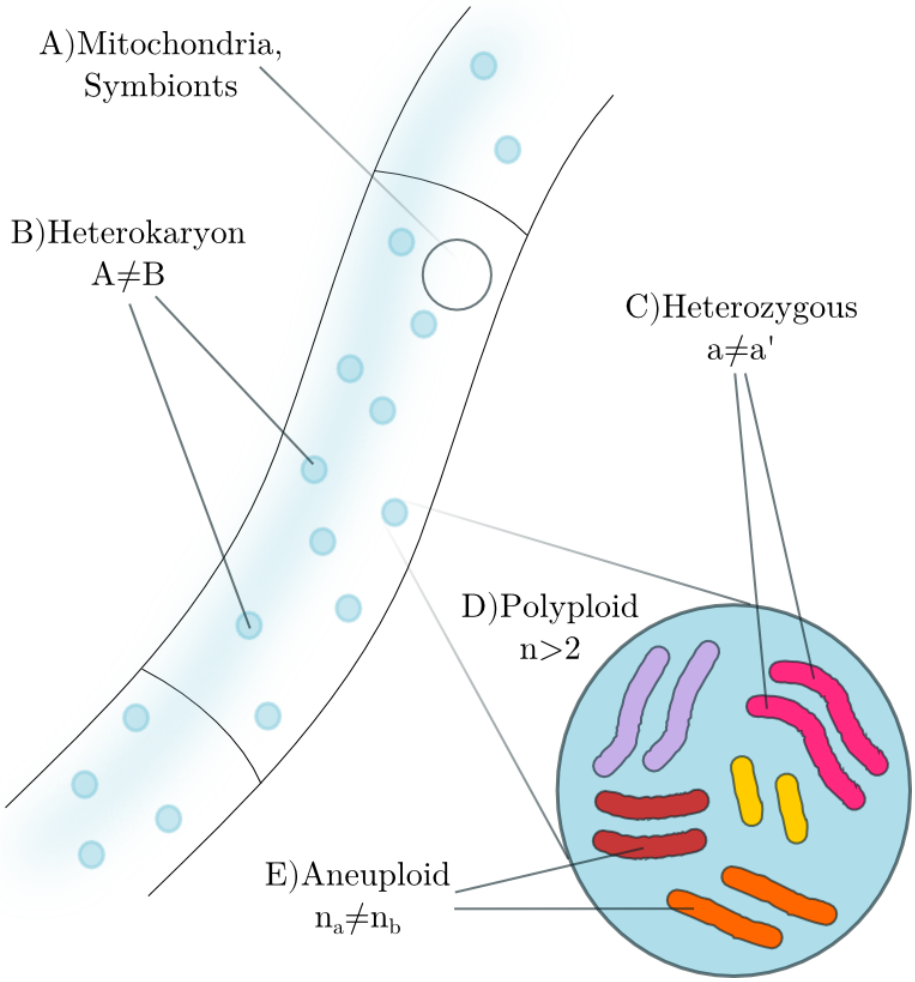

Figure 1: Schematic representation of the different deviations from standard genome structure in fungi.

A) Many fungi harbor symbiotic associations with bacteria or other eukaryotes whose DNA will be sequenced alongside the fungal DNA. **B)** Filamentous fungi might present populations of nuclei that present sequence differences among them, a phenomenon known as heterokaryosis. **C)** Sequence differences between recombining chromosomes are possible, known as heterozygosis. **D)** Most eukaryotes present a number of chromosome copies equal to either one (haploid) or two. Numbers above 2 are known as polyploidies. **E)** Chromosome number might vary between chromosomes. These deviations are known as aneuploidies. All these phenomena carry important biological consequences and might occur at the same time, difficulting genome assembly and analyses.

3.4 Hybridization in Fungi

Allopolyploidy emerges after two genetically separated lineages mate and form a hybrid that duplicates its genome. Hybrid genomes tend to be unstable, since the chromosomic structure of both parentals might not be equivalent, preventing recombination (Bennett *et al.* , 2014; Kumaran *et al.* , 2013; Li *et al.* , 2012; Pryszcz *et al.* , 2015). By doubling the chromosomes, the genome becomes stable again, since each chromosome will be able to recombine with its own copy. The hybrid usually presents intermediate phenotypes compared to the parentals, although novel phenotypes might arise, a phenomenon commonly known as "hybrid vigor" (Bennett *et al.* , 2014; Otto, 2007; Schoenfelder & Fox, 2015). Hybridization has been historically an important subject of interest for botanist, since plants can be crossed easily and present plenty of visually recognizable traits. If plants, limited as they are by their complex nature, are able to interbreed with such freedom, it is reasonable to expect that fungi should potentially have fewer limits in their ability to form hybrids. In practice though, the lack of distinctive morphological characteristics prevent mycologists to recognize hybrids by traditional means. Only recently, with the advent of genomics, the widespread existence of fungal hybrids has started to be noted. It is important to note that hybrid research has traditionally focused on diploid organisms, while many fungi are haploid. This does not impede allopolyploidization. In fact, mating between two highly divergent genomes could produce a hybrid that would have problems to perform meiotic reduction. By skipping this process, just like a regular allotetraploid, you would obtain an "allodiploid" that would have two complete sets of sub-genomes that would not suffer recombination (Campbell *et al.* , 2016). Additionally, fungal hybrids very rarely keep a balanced genetic contribution form their parental strains, further challenging detection of these phenomena. A mixture of selection and genomic instability tend to make hybrid genomes a sort of a "collage" (Bennett *et al.* , 2014; Borneman *et al.* , 2014; Möller & Stukenbrock, 2017; Pryszcz *et al.* , 2015). The overall hybrid genome tends to be very heterozygous since includes homologous regions from two distant parentals. Some regions, commonly referred as loss of heterozygosity, present very low polymorphism density and tend to present higher identity with one of the parentals (Berman, 2016; Ga *et al.* , 2013; Pryszcz *et al.* , 2015). These regions might affect just a few thousands of bases or comprise entire chromosomes. The first case is caused by true meiotic recombination and is a powerful tool to identify positively selected genes that might correlate with ecological adaptations. The second case is usually caused by aneuploidies and other drastic rearrangements, for instance by loss of one of the parental chromosomes and subsequent duplication of the remaining one, or *vice versa*.

The success of *Saccharomyces cerevisiae* as a research model in genetics and biochemistry led to the identification of its sexual cycle and of metabolic traits that could be used as makers for different lineages and their progeny. This allowed the discovery that several *Saccharomyces* strains, classified as independent species, were in fact hybrids (Borneman *et al.* , 2011; Dujon, 2010; Hittinger, 2013; Kumaran *et al.* , 2013; Leducq *et al.* , 2016; Morales & Dujon, 2012; Walther *et al.* , 2014). The advent of genome sequencing was a revolution for yeast research, and showed that *Saccharomyces* ability to hybridize was not exceptional within Saccharomycotina. Hybrid yeasts seem to be common in industrial environments and some of them are particularly relevant as main fermenters in food and beverage industry (Borneman *et al.* , 2014; Hellborg & Piskur, 2009; James *et al.* , 2005; Morales & Dujon, 2012; Walther *et al.* , 2014). Hybrid yeasts can also be found within clinical isolates (Gabaldón *et al.* , 2016; Mixão & Gabaldón, 2018). In this regard, the *Candida parapsilosis* species complex is a particularly interesting example, with numerous described hybrid isolates that apparently present higher pathogenic capabilities than their non hybrid relatives (Gabaldón *et al.* , 2016; Mixão & Gabaldón, 2018; Pryszcz *et al.* , 2014,1). Also, the inability to find the parental strains in some hybrids might be caused by the same hybrid overtaken the ecological niche of the parental and causing its demise (Depotter *et al.* , 2016; Pryszcz *et al.* , 2015). All these cases suggest that hybridization is a powerful driver for adaptation to novel environments, including new hosts. However, it is important to note that we still know very little about the physiology of yeasts in natural environments, and while this hypothesis is indeed attractive, it is currently very difficult to demonstrate. We also have evidence of an ancient polyploidization event that occurred 100Mya in the Saccharomycetaceae (Saccharomycotina), affecting the common ancestor of the genera *Saccharomyces*, *Nakaseomyces*, *Kazachstania* and *Naumovozyma* (Hittinger, 2013;

Marcet-houben & Gabaldón, 2015; Wolfe & Shields, 1997).

Hybridization outside Saccharomycotina has been described, although sampling in other groups is certainly not as extensive. The *Cryptococcus neoformans* species complex contains several hybrids between distantly related strains of (Aminnejad *et al.* , 2012; Cogliati *et al.* , 2012; Li *et al.* , 2012). In the case of the so called AD serotypes, one of the parental species seems to be geographically restricted to certain areas in Africa, the hybrid has spread all over the world, again suggesting that hybridization might have provided the pathogen with some advantage (Cogliati *et al.* , 2012; Li *et al.* , 2012). Mating between *C. neoformans* and *C. gatti* is possible in laboratory conditions, producing a viable offspring that nonetheless possesses a highly unstable genome (Aminnejad *et al.* , 2012). The hybrid rapidly loses and rearranges chromosomes in a manner similar to the parasexual cycle of *Candida albicans*, suggesting that hybridization could provide rapid adaptation by generating highly volatile genomic configurations (Aminnejad *et al.* , 2012; Forche, 2012; Morrow & Fraser, 2013). Population analysis in *Coccidioides* (Eurotiomycetes), another genus comprising human pathogens, detected recent hybridization (Neafsey *et al.* , 2010). Hybridization is common in the grass endophyte *Epichlöe/Neotyphodium* (Sordariomycetes) (Hamilton *et al.* , 2009; Saari & Faeth, 2012; Shoji *et al.* , 2015), and it has been shown that certain events enhance the ability to colonize the grass under stressful conditions. Numerous examples of hybridization events, both *in vitro* and in natural populations have been described for plant pathogens in the Basidiomycota and the Ascomycota (Park & Wellings, 2012; Sriswasdi *et al.* , 2016; Stukenbrock, 2016; Stukenbrock *et al.* , 2012). In this regard, introduction of crops in new territories, global trade and movement of people around the globe provide routes for the contact between otherwise geographically isolated populations and may potentially produce novel hybrid strains that might turn into emergent pathogens (Gonthier *et al.* , 2004; Stukenbrock, 2016; Stukenbrock *et al.* , 2011). Beyond the already mentioned whole genome duplication event in the Saccharomycetaceae, the order Mucorales (Mucoromycotina) seems to have experienced at least two well characterized events of this kind: An ancient event affecting the ancestor of *Mucor* and *Phycomyces* (Corrochano *et al.* , 2016), and a more recent one within the genus *Rhizopus* (Ma *et al.* , 2009). The apparent scarcity of ancient polyploidization in fungi, as compared to animals or plants, is at odds with their expected higher plasticity and is likely the result of a higher difficulty of detecting them (Campbell *et al.* , 2016). Finally, an additional whole genome duplication event has been described for the hyperhalotolerant black yeast *Hortaea werneckii* (Lenassi *et al.* , 2013; Sinha *et al.* , 2017). This event has contributed to the expansion on cationic transporters in *Hortaea*, an important arsenal to survive high salinity. We are confident that many ancient whole genome duplication events are yet to be identified and will be discovered once sufficient genome sampling is available for enough fungal clades. Very likely, at least some of these events will correlate with the emergence of important traits and ecological transitions within the group, just like it has been observed for plants and metazoans.

Unlike plants, for which many ancient polyploidization events have been identified, very few have been identified for fungi (Campbell *et al.* , 2016). Fungi possess much faster evolutionary rates, higher chromosome plasticity and higher pressure for keeping a reduced genome while having usually less freedom in this regard due to a higher genome compaction. Cytogenetic studies in fungi are much less prevalent than in plants and animals, too, requiring more costly approaches such as pulse-field electrophoresis. All this factors reduce the detectability of signals that have been used in plants (Carretero-Paulet *et al.* , 2015; Del Pozo & Ramirez-Parra, 2015; Jiao *et al.* , 2011; Leitch & Leitch, 2012; Soltis *et al.* , 2015) and animals (Dehal & Boore, 2005; Kenny *et al.* , 2016; Li *et al.* , 2018b; Schwager *et al.* , 2017; Van de Peer, 2004; Van de Peer *et al.* , 2003) to identify ancient polyploidies, such as synteny, chromosome numbers or widespread and phylogenetically restricted gene duplication. In short evolutionary terms, chromosome aberrations present some other emergent ecological properties. Even in highly homogeneous environments, such as liquid laboratory cultures, small microniches that impose differential selective pressures might emerge (Ibarra *et al.* , 2002; Rosenzweig *et al.* , 1994; Wortel *et al.* , 2016). If that is the case, two or more reversible chromosomic states might coexist within a population, each one employing a different strategy. Polyploidies might produce meiotic and mitotic impairments, and imply higher nitrogen and phosphorus costs per cell division, which in turn slows growth under optimal conditions (Otto, 2007; Schoenfelder & Fox, 2015; Scott *et al.* , 2017). It is important

to note that such conditions are usually met almost exclusively in laboratory settings. It is then reasonable to assume that putative polyploid or aneuploid that have been growth in axenic cultures for long enough could streamline their genome towards an ideal simple haploid or diploid state that maximizes potential growth rate. Non-canonical chromosomic conformations in environmental fungi might even prevent growth under laboratory conditions, thus contributing to the great plate count anomaly. Even more, this could explain why culture-based diagnostic methods of fungal infections are so prone to negative results (Ostrosky-zeichner, 2012). Genome sequencing projects tend to focus on reference strains or on strains that a particular laboratory uses as a model. In both cases, it is very likely that these strains have been grown for years in non-limiting conditions. Even in those cases in which an isolate with any of these chromosome aberrations is sequenced, these same genomic circumstances require specialized experimental and computational approaches that are far from standard, which require expertise and increased costs. Polyploid strains, specially allopolyploids, produce highly fragmented assemblies that often present an inflated genome size, due to their high heterozygosity (Huang *et al.* , 2017; Kajitani *et al.* , 2014; Pryszcz & Gabaldón, 2016; Safonova *et al.* , 2015). The same can be said for supernumerary chromosomes, which might pass unnoticed as a collection of highly fragmented scaffolds within an otherwise typical assembly. Because of this, many hybrids, polyploids and aneuploids might have been already sequenced but still not described. Direct measurements of genome size by flow cytometry are a poor man's substitution and genome size comparisons among related species or strains provide a good initial approach to the problem. This strategy is not reliable enough to identify aneuploidies, but at least has the advantage to distinguish between allopolyploids and heterokaryons.

3.5 Evolutionary implications of the heterokaryon

The thallus of most fungal species forms a syncytium, a cell with several nuclei that share a cytoplasm. The amount of nuclei in a particular fungal colony can easily be in the order of thousands (Roper *et al.* , 2012,1) and thus the assumption that all nuclei are genotypically identical is an oversimplification. The coexistence of two or more genetically distinct nuclear populations within a syncytium is called heterokaryosis. If the phenotypical response of these different populations of nuclei are different, variations in the proportions of each nuclear population can translate to phenotypical variation in the whole colony. This was proposed and demonstrated by Jinks on the base of experiments using wild heterokaryotic *Penicillium* (Jinks, 1952; Strom & Bushley, 2016). In nature, fungi might cover great surfaces without breaking cytoplasm continuity, with astounding examples such as the genus *Armillaria* (Sipos *et al.* , 2018). Heterokaryosis would then affect local responses to stimuli within different areas of the same mycelium (Jany & Pawlowska, 2010; Roper *et al.* , 2012).

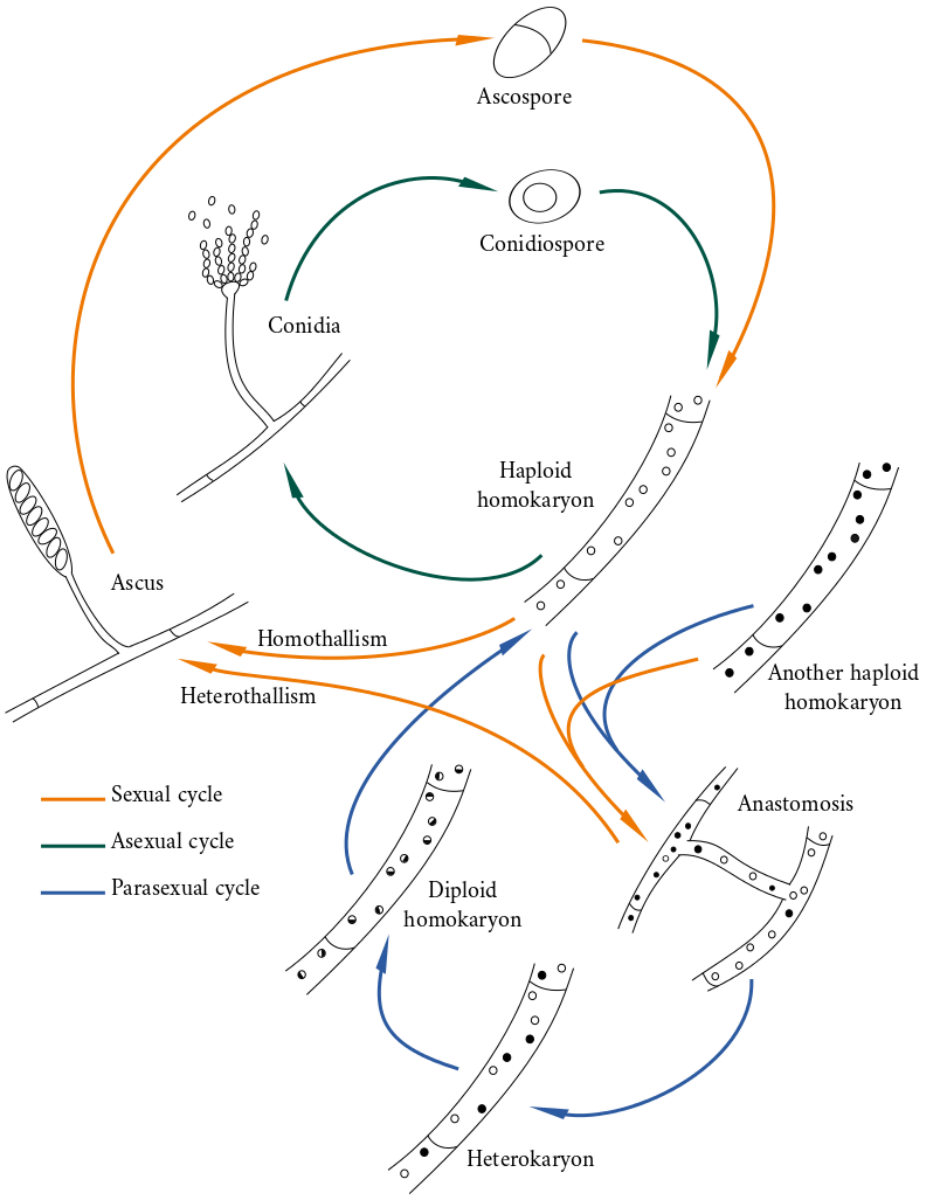

Figure 2: Life cycle of a generic Pezizomycotina.

In most cases, vegetative thallus is formed by a homokaryotic mycelium. **Sexual cycle**: Homokaryotic mycelium might transform into ascoid mycelium, where nuclei undergo meiosis and form ascospores in asci. This change might occur after fusion with another compatible mycelium, producing recombination between two strains (heterothallism); or occur in the nuclear population within the mycelium, without genetic exchange (homothallism). **Asexual cycle**: Vegetative mycelium can form conidia that produce conidiospores by simple mitosis. **Parasexual cycle**: Two compatible vegetative mycelia might fuse through anastomosis, interchanging nuclei and forming an haploid heterokaryon. At some point, the different nuclei might fuse, forming a diploid homokaryon that undergoes concerted chromosome loss to regain a haploid state in the absence of meiosis. This process returns the mycelium to a haploid homokaryon state.

At least theoretically, heterokaryons are expected to be unstable (Hallatschek & Nelson, 2008; Roper *et al.* , 2012,1). If nuclear populations spread based on their relative fitness and simple diffusion, one of the populations should be out-competed by the other or disappear due to stochastic effects; not unlike alleles within any genetic population. Furthermore, “nuclear death” in filamen-

tous fungi has been described, as nuclei from senescent mycelia can enter apoptosis and recycle their nutrients (Maheshwari, 2005). This implies that nuclei with low fitness will not just dilute within the population if is is big enough, but they will be systematically eliminated. Indeed, phenotypic heterogeneity in nuclei sharing a cytoplasm can be better understood under the light of population dynamics. For example under the right conditions, such as growth in supplemented media, mutant nuclei can out-compete the wild phenotype (Maheshwari, 2005; Ryan & Lederberg, 1946). On the other hand, nuclei carrying different mutations can also complement each other, as proved for the carotenoid biosynthetic pathway in *Phycomyces* (De la Guardia *et al.* , 1971; Sanz *et al.* , 2002; Strom & Bushley, 2016). Based on this, mixed nuclear populations with distinct genetic backgrounds might, in theory, become stable under so called "Black Queen" scenarios (Morris, 2015). That is, in a simple community of two members (A and B), if A loses the ability to perform certain task that can be fulfilled sufficiently by B, then B will be "trapped" and unable to lose that function, as this would mean the collapse of the community. If the same situation happens for another function, but this time B is the one losing its gene, A and B would be "reciprocal parasites", needing each other for survival. This situation was artificially generated in a classic experiment with *Neurospora* (Beadle & Coonradt, 1944; Strom & Bushley, 2016). The population nature of the heterokaryon adds a new layer of phenotypical complexity without developing complex regulatory mechanisms (Anderson *et al.* , 2013; Dundon *et al.* , 2016; Maheshwari, 2005; Roper *et al.* , 2012,1,1; Strom & Bushley, 2016) or increasing effective genome size, since each nucleus contains roughly the same genetic information. Experimental evidence in *Neurospora tetrasperma* suggests that, at least in this species, nuclear populations are kept at controlled proportions that vary depending on the life stage of the fungus (Johannesson & Samils, 2014; Roper *et al.* , 2012,1,1). Despite the astounding growth speed of this mould (Ryan *et al.* , 1943) and its asynchronous nuclear division, the heterokaryon is stable over long periods of time. It is important to note that nuclear division in *Neurospora* is not restricted to hyphal tips, and active cytoplasmic currents provide the growing tip with fresh nuclei generated throughout the whole colony (Maheshwari, 2005; Roper *et al.* , 2012,1). Cytoplasmic currents are responsible for preventing stochastic extinction by actively mixing the cytoplasm of the whole colony (Johannesson & Samils, 2014; Roper *et al.* , 2013,1). In *Eremothecium*, however, nuclear migration is mediated by the cytoskeleton (Anderson *et al.* , 2013; Dundon *et al.* , 2016; Gibeaux *et al.* , 2017; Gladfelter, 2006). The ability to export nuclei from far regions if the network would provide a steady supply of nuclei and help keeping directed growth even if comparatively slower nuclei are present, as is the case for those carrying chromosome aberrations that, as mentioned before, probably populate in some degree any population (Anderson *et al.* , 2015; Nobre *et al.* , 2014). It is important to mention here that nuclei seem to control cytoplasmic territories with considerable autonomy from the rest, also known as "cells within cells" (Anderson *et al.* , 2013; Nair *et al.* , 2010; Roberts & Gladfelter, 2015; Roper *et al.* , 2012,1). Old protoplasm fusion experiments force this autonomy to the point of achieving that nuclei from even different fungal phyla share a cytoplasm (Kavanagh & Whittaker, 1996; Peberdy, 1979,8; Strom & Bushley, 2016).

Filamentous fungi can fuse their hyphae during their normal sexual cycle, and formation of dikaryotic hyphae is a defining trait of Dikarya. If the parentals are too divergent these mixtures might become trapped, unable to undergo meiosis. Even then, inter-species heterokaryon might arise and be stable in nature. Using sequencing data present similar challenges in these cases as with hybrid genomes, and if such approach is exclusively taken it might be difficult to distinguish both scenarios. The existence of short haploidized regions in an apparent diploid genome is suggestive of the existence of two non-recombining sub-genomes and compatible with heterokaryosis. Also, significant deviations in the expected even proportions on reference versus variation SNPs might be used to identify heterokaryotic genomes in some cases. Successive formation of heterokaryons might allow the fungus to "update" its genome on the fly, fusing its cytoplasm with new individuals as conditions change (Beadle & Coonradt, 1944; Strom & Bushley, 2016). However this carries the risk of being overtaken by a faster dividing nuclear population, as well as spread of viruses and other infectious elements (Aanen *et al.* , 2008; Johannesson & Samils, 2014; Saupe, 2000; Strom & Bushley, 2016). For these reasons, fungi present genetic mechanisms to prevent inter-species fusion, although these barriers are not impenetrable (Daskalov *et al.* , 2017; Ishikawa *et al.* , 2012; Saupe, 2000; Van der Nest *et al.* , 2014). Even if temporary, exotic unions might have long lasting effects by promoting genome rearrangements and transference of genetic material (James *et al.* ,

2008; Kinsey, 1990; Soanes & Richards, 2014; Van Der Does & Rep, 2012). Heterokaryosis might emerge spontaneously without the need of mating or high heterozygosity, for instance through ploidy changes (Anderson *et al.* , 2015). This is the case of the filamentous Saccharomycotina *Eremothecium gossypii* (syn. *Ashbya gossypii*). *E. gossypii* presents hyphae whose nuclei divide independently (Anderson *et al.* , 2013; Dundon *et al.* , 2016; Gladfelter, 2006; Nair *et al.* , 2010) and form populations with varying karyotype. The proportion of nuclei carrying abnormal chromosome numbers varies under stressful conditions and mycelial age (Fisher *et al.* , 2012). Even more, under environmental conditions that positively select chromosomic aberrations, the presence of normal nuclei within the cytoplasm might help buffer deleterious consequences (Anderson *et al.* , 2015). While most of the experimental work has been carried out in filamentous Ascomycota, the dikaryon phase in these organisms is usually short lived. In Basidiomycota, on the other hand, the dikaryon forms most of the vegetative thallus. In most cases, the fungus tightly controls nuclear division by forcing synchronization through clamp connections (Iwasa *et al.* , 1998; Maheshwari, 2005; Raudaskoski & Kothe, 2010; Shepherd *et al.* , 1993). Even then, nuclear competition and population dynamics have been described and studied in *Heterobasidion* (Garbelotto & Gonthier, 2013; Garbelotto *et al.* , 2004; Giordano *et al.* , 2018; James *et al.* , 2008,0) and *Termitomyces* (Nobre *et al.* , 2014). Mycorrhizic fungi also seem to present high heterokaryosis in nature (Manning & Callaghan, 2008; Manning *et al.* , 2007; Wyss *et al.* , 2016), forming nuclear populations that are maintained over time through highly multinucleated spores (Bever & Wang, 2005; Boon *et al.* , 2015; Chagnon, 2014; Jany & Pawlowska, 2010). However, there is considerable discrepancy between studies and methods and the amount of estimated divergence (Kuo *et al.* , 2014; Lin *et al.* , 2014a; Ropars & Corradi, 2015). In these fungi, nuclear populations vary depending on the nature on the fungal host, and might explain the apparent low specificity on mycorrhizae-plant interactions (Angelard *et al.* , 2014; Chagnon, 2014).

References

Aanen, Duur K., Debets, Alfons J.M., de Visser, J. Arjan G.M., & Hoekstra, Rolf F. 2008. The social evolution of somatic fusion. *BioEssays*, **30**(11-12), 1193–1203.

Akao, Takeshi, Yashiro, Isao, Hosoyama, Akira, Kitagaki, Hiroshi, Horikawa, Hiroshi, Watanabe, Daisuke, Akada, Rinji, Ando, Yoshinori, Harashima, Satoshi, Inoue, Toyohisa, Inoue, Yoshiharu, Kajiwara, Susumu, Kitamoto, Katsuhiko, Kitamoto, Noriyuki, Kobayashi, Osamu, Kuhara, Satoru, Masubuchi, Takashi, Mizoguchi, Haruhiko, Nakao, Yoshihiro, Nakazato, Atsumi, Namise, Masahiro, Oba, Takahiro, Ogata, Tomoo, Ohta, Akinori, Sato, Masahide, Shibasaki, Seiji, Takatsume, Yoshifumi, Tanimoto, Shota, Tsuboi, Hirokazu, Nishimura, Akira, Yoda, Koji, Ishikawa, Takeaki, Iwashita, Kazuhiro, Fujita, Nobuyuki, & Shimoi, Hitoshi. 2011. Whole-genome sequencing of sake yeast Saccharomyces cerevisiae Kyokai no. 7. *DNA research : an international journal for rapid publication of reports on genes and genomes*, **18**(6), 423–34.

Albertin, Warren, & Marullo, Philippe. 2012. Polyploidy in fungi: evolution after whole-genome duplication. *Proceedings. Biological sciences*, **279**(1738), 2497–509.

Aminnejad, Mojgan, Diaz, Mara, Arabatzis, Michael, Castañeda, Elizabeth, Lazera, Marcia, Velegraki, Aristea, Marriott, Deborah, Sorrell, Tania C., & Meyer, Wieland. 2012. Identification of Novel Hybrids Between Cryptococcus neoformans var. grubii VNI and Cryptococcus gattii VGII. *Mycopathologia*, **173**(5-6), 337–346.

Anderson, Cori A, Eser, Umut, Korndorf, Therese, Borsuk, Mark E, Skotheim, Jan M, & Gladfelter, Amy S. 2013. Nuclear Repulsion Enables Division Autonomy in a Single Cytoplasm. *Current Biology*, **23**(20), 1999–2010.

Anderson, Cori A, Roberts, Samantha, Zhang, Huaiying, Kelly, Courtney M, Kendall, Alexxy, Lee, ChangHwan, Gerstenberger, John, Koenig, Aaron B, Kabeche, Ruth, & Gladfelter, Amy S. 2015. Ploidy variation in multinucleate cells changes under stress. *Molecular biology of the cell*, **26**(6), 1129–40.

Anderson, Matthew Z, Saha, Amrita, Haseeb, Abid, & Bennett, Richard J. 2017. A chromosome 4 trisomy contributes to increased fluconazole resistance in a clinical isolate of Candida albicans. *Microbiology (Reading, England)*, **163**(6), 856–865.

Andersson, Jan O. 2005. Lateral gene transfer in eukaryotes. *Cellular and molecular life sciences*, **62**(11), 1182–97.

Andersson, Jan O. 2009. Gene transfer and diversification of microbial eukaryotes. *Annual review of microbiology*, **63**(jan), 177–93.

Angelard, Caroline, Tanner, Colby J., Fontanillas, Pierre, Niculita-Hirzel, Hélène, Masclaux, Frédéric, & Sanders, Ian R. 2014. Rapid genotypic change and plasticity in arbuscular mycorrhizal fungi is caused by a host shift and enhanced by segregation. *ISME Journal*, **8**(2), 284–294.

Armijos-Jaramillo, Vinicio D, Sukno, Serenella A, & Thon, Michael R. 2014. Identification of horizontally transferred genes in the genus Colletotrichum reveals a steady tempo of bacterial to fungal gene transfer. *BMC genomics*, **16**(2), 1–16.

Armitage, Andrew D, Taylor, Andrew, Sobczyk, Maria K, Baxter, Laura, Greenfield, Bethany P.J., Bates, Helen J, Wilson, Fiona, Jackson, Alison C, Ott, Sascha, Harrison, Richard, & Clarkson, John P. 2017a. Pangenomic analysis reveals pathogen-specific regions and novel effector candidates in Fusarium oxysporum f.sp. cepae. *bioRxiv*, sep, 182238.

Avramova, Marta, Cibrario, Alice, Peltier, Emilien, Coton, Monika, Coton, Emmanuel, Schacherer, Joseph, Spano, Giuseppe, Capozzi, Vittorio, Blaiotta, Giuseppe, Salin, Franck, Dols-Lafargue, Marguerite, Grbin, Paul, Curtin, Chris, Albertin, Warren, & Masneuf-Pomarede, Isabelle. 2018. Brettanomyces bruxellensis population survey reveals a diploid-triploid complex structured according to substrate of isolation and geographical distribution. *Scientific reports*, **8**(1), 4136.

Beadle, G W, & Coonradt, V L. 1944. Heterocaryosis in Neurospora Crassa. *Genetics*, **29**(3), 291–308.

Bennett, Richard J, Forche, Anja, & Berman, Judith. 2014. Rapid mechanisms for generating genome diversity: whole ploidy shifts, aneuploidy, and loss of heterozygosity. *Cold Spring Harbor perspectives in medicine*, **4**(10).

Berman, Judith. 2016. Ploidy plasticity: a rapid and reversible strategy for adaptation to stress. *FEMS Yeast Research*, **16**(3), fow020.

Berman, Judith, Wertheimer, Noa Blutraich, & Stone, Neil. 2016. Ploidy dynamics and evolvability in fungi. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **371**(20150461).

Bettini, Priscilla P., Frascella, Arcangela, Kolarik, Miroslav, Comparini, Cecilia, Pepori, Alessia L., Santini, Alberto, Scala, Felice, & Scala, Aniello. 2014. Widespread horizontal transfer of the cerato-ulmin gene between Ophiostoma novo-ulmi and Geosmithia species. *Fungal Biology*, **8**(118), 663–674.

Bever, James D., & Wang, Mei. 2005. Arbuscular mycorrhizal fungi: Hyphal fusion and multigenomic structure. *Nature*, **433**(7022), E3–E4.

Bonney, Megan E, Moriya, Hisao, & Amon, Angelika. 2015. Aneuploid proliferation defects in yeast are not driven by copy number changes of a few dosage-sensitive genes. *Genes & development*, **29**(9), 898–903.

Boon, Eva, Halary, Sébastien, Bapteste, Eric, & Hijri, Mohamed. 2015. Studying genome heterogeneity within the arbuscular mycorrhizal fungal cytoplasm. *Genome biology and evolution*, **7**(2), 505–21.

Borneman, Anthony R., Desany, Brian A., Riches, David, Affourtit, Jason P., Forgan, Angus H., Pretorius, Isak S., Egholm, Michael, & Chambers, Paul J. 2011. Whole-Genome Comparison Reveals Novel Genetic Elements That Characterize the Genome of Industrial Strains of Saccharomyces cerevisiae. *PLoS Genetics*, **7**(2), e1001287.

Borneman, Anthony R, Zeppel, Ryan, Chambers, Paul J, & Curtin, Chris D. 2014. Insights into the Dekkera bruxellensis genomic landscape: comparative genomics reveals variations in ploidy and nutrient utilisation potential amongst wine isolates. *PLoS genetics*, **10**(2), e1004161.

Brankovics, Balázs, Kulik, Tomasz, Sawicki, Jakub, Bilska, Katarzyna, Zhang, Hao, de Hoog, G Sybren, van der Lee, Theo AJ, Waalwijk, Cees, & van Diepeningen, Anne D. 2018. First steps towards mitochondrial pan-genomics: Detailed analysis of Fusarium graminearum mitogenomes. *PeerJ Preprints*, jul.

Brown, Alistair J P, Budge, Susan, Kaloriti, Despoina, Tillmann, Anna, Jacobsen, Mette D, Yin, Zhikang, Ene, Iuliana V, Bohovych, Iryna, Sandai, Doblin, Kastora, Stavroula, Potrykus, Joanna, Ballou, Elizabeth R, Childers, Delma S, Shahana, Shahida, & Leach, Michelle D. 2014. Stress adaptation in a pathogenic fungus. *The journal of Experimental Biology*, **217**, 144–155.

Campbell, Matthew A., Ganley, Austen R. D., Gabaldón, Toni, & Cox, Murray P. 2016. The Case of the Missing Ancient Fungal Polyploids. *The American Naturalist*, **188**(6), 602–614.

Carretero-Paulet, Lorenzo, Chang, Tien Hao, Librado, Pablo, Ibarra-Laclette, Enrique, Herrera-Estrella, Luis, Rozas, Julio, & Albert, Victor A. 2015. Genome-wide analysis of adaptive molecular evolution in the carnivorous plant utricularia gibba. *Genome Biology and Evolution*, **7**(2), 444–456.

Chagnon, Pierre-Luc. 2014. Ecological and evolutionary implications of hyphal anastomosis in arbuscular mycorrhizal fungi. *FEMS Microbiology Ecology*, **88**, 437–444.

Chen, Eric C. H., Morin, Emmanuelle, Beaudet, Denis, Noel, Jessica, Yildirir, Gokalp, Ndikumana, Steve, Charron, Philippe, St-Onge, Camille, Giorgi, John, Krüger, Manuela, Marton, Timea, Ropars, Jeanne, Grigoriev, Igor V., Hainaut, Matthieu, Henrissat, Bernard, Roux, Christophe, Martin, Francis, & Corradi, Nicolas. 2018a. High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont Rhizophagus irregularis. *New Phytologist*, jan.

Chiara, Matteo, Fanelli, Francesca, Mulè, Giuseppina, Logrieco, Antonio F., Pesole, Graziano, Leslie, John F., Horner, David S., & Toomajian, Christopher. 2015. Genome Sequencing of Multiple Isolates Highlights Subtelomeric Genomic Diversity within Fusarium fujikuroi. *Genome Biology and Evolution*, **7**(11), 3062–3069.

Cogliati, M., Barchiesi, F., Spreghini, E., & Tortorano, Anna Maria. 2012. Heterozygosis and Pathogenicity of Cryptococcus neoformans AD-Hybrid Isolates. *Mycopathologia*, **173**(5-6), 347–357.

Conant, Gavin C., & Wolfe, Kenneth H. 2008. Turning a hobby into a job: How duplicated genes find new functions. *Nature Reviews Genetics*, **9**(12), 938–950.

Corrochano, Luis M, Kuo, Alan, Marcet-Houben, Marina, Polaino, Silvia, Salamov, Asaf, Villalobos-Escobedo, José M, Grimwood, Jane, Álvarez, M Isabel, Avalos, Javier, Bauer, Diane, Benito, Ernesto P, Benoit, Isabelle, Burger, Gertraud, Camino, Lola P, Cánovas, David, Cerdá-Olmedo, Enrique, Cheng, Jan-Fang, Domínguez, Angel, Eliáš, Marek, Eslava, Arturo P, Glaser, Fabian, Gutiérrez, Gabriel, Heitman, Joseph, Henrissat, Bernard, Iturriaga, Enrique A, Lang, B Franz, Lavín, José L, Lee, Soo Chan, Li, Wenjun, Lindquist, Erika, López-García, Sergio, Luque, Eva M, Marcos, Ana T, Martin, Joel, McCluskey, Kevin, Medina, Humberto R, Miralles-Durán, Alejandro, Miyazaki, Atsushi, Muñoz-Torres, Elisa, Oguiza, José A, Ohm, Robin A, Olmedo, María, Orejas, Margarita, Ortiz-Castellanos, Lucila, Pisabarro, Antonio G, Rodríguez-Romero, Julio, Ruiz-Herrera, José, Ruiz-Vázquez, Rosa, Sanz, Catalina, Schackwitz, Wendy, Shahriari, Mahdi, Shelest, Ekaterina, Silva-Franco, Fátima, Soanes, Darren, Syed, Khajamohiddin, Tagua, Víctor G, Talbot, Nicholas J, Thon, Michael R, Tice, Hope, de Vries, Ronald P, Wiebenga, Ad, Yadav, Jagjit S, Braun, Edward L, Baker, Scott E, Garre, Victoriano, Schmutz, Jeremy, Horwitz, Benjamin A, Torres-Martínez, Santiago, Idnurm, Alexander, Herrera-Estrella, Alfredo, Gabaldón, Toni, & Grigoriev, Igor V. 2016. Expansion of Signal Transduction Pathways in Fungi by Extensive Genome Duplication. *Current biology*, **26**(12), 1577–1584.

Daskalov, Asen, Heller, Jens, Herzog, Stephanie, Fleissner, André, & Glass, N. Louise. 2017. Molecular Mechanisms Regulating Cell Fusion and Heterokaryon Formation in Filamentous Fungi. *Microbiology Spectrum*, **5**(2).

De la Guardia, M D, Aragón, C M, Murillo, F J, & Cerdá-Olmedo, E. 1971. A carotenogenic enzyme aggregate in Phycomyces: evidence from quantitive complementation. *Proceedings of the National Academy of Sciences*, **68**(9), 2012–5.

de Vries, Ronald P, Riley, Robert, Wiebenga, Ad, Aguilar-Osorio, Guillermo, Amillis, Sotiris, Akemi Uchima, Cristiane, Anderluh, Gregor, Asadollahi, Mojtaba, Askin, Marion, Barry, Kerrie, Battaglia, Evy, Bayram, Özgür, Benocci, Tiziano, Braus-Stromeyer, Susanna A, Caldana, Camila, Cánovas, David, Cerqueira, Gustavo C, Chen, Fusheng, Chen, Wanping, Choi, Cindy, Clum, Alicia, Augusto Corrêa dos Santos, Renato, Karaffa, Levente, Karányi, Zsolt, Kraševec, Nada, Kuo, Alan, Kusch, Harald, LaButti, Kurt, Lagendijk, Ellen L, Lapidus, Alla, Levasseur, Anthony, Lindquist, Erika, Lipzen, Anna, Mulé, Giuseppina, Yee Ngan, Chew, Orejas, Margarita, Orosz, Erzsébet, Paul Ouedraogo, Jean, Overkamp, Karin M, Park, Hee-Soo, Perrone, Giancarlo, Piumi, Francois, Punt, Peter J, J Ram, Arthur F, Ramón, Ana, Rauscher, Stefan, Record, Eric, Mauricio Riaño-Pachón, Diego, Robert, Vincent, Röhrig, Julian, Ruller, Roberto, Salamov, Asaf,

Salih, Nadhira S, Samson, Rob A, Sándor, Erzsébet, Sanguinetti, Manuel, Schütze, Tabea, Sepčić, Kristina, Shelest, Ekaterina, Sherlock, Gavin, Sophianopoulou, Vicky, Squina, Fabio M, Sun, Hui, Susca, Antonia, Todd, Richard B, Tsang, Adrian, Unkles, Shiela E, van de Wiele, Nathalie, van Rossen-Uffink, Diana, Velasco de Castro Oliveira, Juliana, Vesth, Tammi C, Visser, Jaap, Yu, Jae-Hyuk, Zhou, Miaomiao, Andersen, Mikael R, Archer, David B, Baker, Scott E, Benoit, Isabelle, Brakhage, Axel A, Braus, Gerhard H, Fischer, Reinhard, Frisvad, Jens C, Goldman, Gustavo H, Houbraken, Jos, Oakley, Berl, Pócsi, István, Scazzocchio, Claudio, Seiboth, Bernhard, VanKuyk, Patricia A, Wortman, Jennifer, Dyer, Paul S, & Grigoriev, Igor V. 2017. Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus Aspergillus. *María Laura Harispe*, **10**, 28.

Dehal, Paramvir, & Boore, Jeffrey L. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS biology*, **3**(10), e314.

Del Pozo, Juan Carlos, & Ramirez-Parra, Elena. 2015. Whole genome duplications in plants: an overview from Arabidopsis. *Journal of Experimental Botany*, **66**(22), 6991–7003.

Depotter, Jasper Rl, Seidl, Michael F, Wood, Thomas A, & Thomma, Bart Phj. 2016. Interspecific hybridization impacts host range and pathogenicity of filamentous microbes. *Current Opinion in Microbiology*, **32**, 7–13.

Dodgson, Stacie E, Kim, Sharon, Costanzo, Michael, Baryshnikova, Anastasia, Morse, Darcy L, Kaiser, Chris A, Boone, Charles, & Amon, Angelika. 2016. Chromosome-Specific and Global Effects of Aneuploidy in Saccharomyces cerevisiae. *Genetics*, **202**(4), 1395–409.

Dong, Suomeng, Raffaele, Sylvain, & Kamoun, Sophien. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. *Current Opinion in Genetics & Development*, **35**(dec), 57–65.

Dujon, Bernard. 2010. Yeast evolutionary genomics. *Nature Publishing Group*, **11**(7), 512–524.

Dujon, Bernard A., & Louis, Edward J. 2017. Genome Diversity and Evolution in the Budding Yeasts (Saccharomycotina). *Genetics*, **206**(2), 717–750.

Dundon, Samantha E R, Chang, Shyr-Shea, Kumar, Abhishek, Occhipinti, Patricia, Shroff, Hari, Roper, Marcus, & Gladfelter, Amy S. 2016. Clustered nuclei maintain autonomy and nucleocytoplasmic ratio control in a syncytium. *Molecular biology of the cell*, **27**(13), 2000–7.

Engel, Stacia R, & Cherry, J Michael. 2013. The new modern era of yeast genomics: community sequencing and the resulting annotation of multiple Saccharomyces cerevisiae strains at the Saccharomyces Genome Database. *Database : the journal of biological databases and curation*, **2013**, bat012.

Farrer, Rhys A, Henk, Daniel A, Garner, Trenton W J, Balloux, Francois, Woodhams, Douglas C, & Fisher, Matthew C. 2013. Chromosomal copy number variation, selection and uneven rates of recombination reveal cryptic genome diversity linked to pathogenicity. *PLoS genetics*, **9**(8), e1003703.

Fisher, Matthew C, Henk, Daniel A, Briggs, Cheryl J, Brownstein, John S, Madoff, Lawrence C, McCraw, Sarah L, & Gurr, Sarah J. 2012. Emerging fungal threats to animal, plant and ecosystem health. *Nature*, **484**(7393), 186–94.

Forche, Anja. 2012. Large-Scale Chromosomal Changes and Associated Fitness Consequences in Pathogenic Fungi. *Curr Fungal Infect Rep*, **29**(6), 997–1003.

Forche, Anja, Alby, Kevin, Schaefer, Dana, Johnson, Alexander D., Berman, Judith, & Bennett, Richard J. 2008. The parasexual cycle in Candida albicans provides an alternative pathway to meiosis for the formation of recombinant strains. *PLoS Biology*, **6**(5), 1084–1097.

Friesen, Timothy L, Stukenbrock, Eva H, Liu, Zhaohui, Meinhardt, Steven, Ling, Hua, Faris, Justin D, Rasmussen, Jack B, Solomon, Peter S, McDonald, Bruce A, & Oliver, Richard P. 2006. Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature Genetics*, **38**(8), 953–956.

Ga, Attila, Gabaldo, Toni, Pryszcz, Leszek P, Ne, Tibor, James, Stephen a., Bond, Christopher J., Stratford, Malcolm, Roberts, Ian N., Louis, V. L., Despons, L., Friedrich, a., Martin, T., Durrens, P., Casaregola, Serge, Neuveglise, C., Fairhead, C., Marck, C., Cruz, J. a., Straub, M.-L., Kugler, V., Sacerdot, Christine, Uzunov, Z., Thierry, a., Weiss, Stéphanie, Bleykasten, C., De Montigny, J., Jacques, Noémie, Jung, P., Lemaire, M., Mallet, Sandrine, Morel, G., Richard, G.-F., Sarkar, a., Savel, G., Schacherer, J., Seret, M.-L., Talla, E., Samson, G., Jubin, C., Poulain, J., Vacherie, B., Barbe, V., Pelletier, E., Sherman, D. J., Westhof, E., Weissenbach, J., Baret, P. V., Wincker, P., Gaillardin, C., Dujon, Bernard, Souciet, J.-L., Marinoni, Gaëlle, Lachance, Marc André, Idnurm, Alexander, Roper, Marcus, Ellison, Christopher E, Taylor, John W, Glass, N Louise, Kumaran, Rajaraman, Yang, Shi Yow, Leu, Jun Yi, Whelan, William L., Markie, David M., Simpkin, K. Graham, Poulter, Russell M., Vasan, Soumini, Deem, Angela, Ramakrishnan, Sreejith, Argueso, Juan Lucas, Malkova, Anna, Kravets, Anatoliy, Yang, Feng, Bethlendy, Gabor, Sherman, Fred, Rustchenko, Elena, Mallet, Sandrine, Weiss, Stéphanie, Jacques, Noémie, Leh-Louis, Véronique, Sacerdot, Christine, Casaregola, Serge, Roper, Marcus, Simonin, Anna, Hickey, Patrick C, Leeder, Abby, Glass, N Louise, Forche, Anja, Alby, Kevin, Schaefer, Dana, Johnson, Alexander D., Berman, Judith, Bennett, Richard J., Ni, Min, Feretzaki, Marianna, Sun, Sheng, Wang, Xuying, Heitman, Joseph, Forche, Anja, Heitman, Joseph, Carter, Dee a, Dyer, Paul S, Soll, David R, Kwon-chung, Kyung J, Fraser, James a., Tamara, L, Roemer, Terry, Krysan, Damian J, White, Theodore C, Findley, Keisha, Thomas, L, Sun, Sheng, Billmyre, R. Blake, Mieczkowski, Piotr a., Heitman, Joseph, Diogo, Dorothée, Bouchier, Christiane, D'Enfert, Christophe, Bougnoux, Marie Elisabeth, Pelin, Adrian, Selman, Mohammed, Aris-Brosou, Stéphane, Farinelli, Laurent, Corradi, Nicolas, Otto, Sarah P., Chakraborty, Uttara, Mohamed, Aiyaz, Kakade, Pallavi, Mugasimangalam, Raja C., Sadhale, Parag P., Sanyala, Kaustuv, Marcet-houben, Marina, Gabaldón, Toni, Butler, Geraldine, Rasmussen, Matthew D, Lin, Michael F, Santos, Maria C Manuel a S, Sakthikumar, Sharadha, Munro, Carol a, Rheinbay, Esther, Grabherr, Manfred, Forche, Anja,

Reedy, Jennifer L, Agrafioti, Ino, Arnaud, Martha B, Bates, Steven, Brown, Alistair J P, Brunke, Sascha, Costanzo, Maria C, Fitzpatrick, David a, de Groot, Piet W J, Harris, David, Hoyer, Lois L, Hube, Bernhard, Klis, Frans M, Kodira, Chinnappa, Lennard, Nicola, Logue, Mary E, Martin, Ronny, Neiman, Aaron M, Nikolaou, Elissavet, Quail, Michael a, Quinn, Janet, Santos, Maria C Manuel a S, Schmitzberger, Florian F, Sherlock, Gavin, Shah, Prachi, Silverstein, Kevin a T, Skrzypek, Marek S, Soll, David R, Staggs, Rodney, Stansfield, Ian, Stumpf, Michael P H, Sudbery, Peter E, Srikantha, Thyagarajan, Zeng, Qiandong, Berman, Judith, Berriman, Matthew, Heitman, Joseph, Gow, Neil a R, Lorenz, Michael C, Birren, Bruce W, Kellis, Manolis, Cuomo, Christina a, Zeyl, Clifford, Bennett, Richard J., Taylor, John W, Hann-soden, Christopher, Branco, Sara, Sylvain, Iman, Ellison, Christopher E, America, South, Peraza-Reyes, Leonardo, Berteaux-Lecellier, Véronique, Rev, Nat, Author, Microbiol, Morrow, Carl a., Fraser, James a., Flagfeldt, Dongmei Bai, Siewers, Verena, Huang, Le, Nielsen, Jens, Morales, Lucia, Dujon, Bernard, Alexander, William G, Peris, David, Pfannenstiel, Brandon T, Opulente, Dana A, Kuang, Meihua, Hittinger, Chris Todd, Anderson, Cori A, Eser, Umut, Korndorf, Therese, Borsuk, Mark E, Skotheim, Jan M, Gladfelter, Amy S, Bertier, Lien, Leus, Leen, Cock, Arthur W A M De, Johannesson, Hanna, Samils, Nicklas, Nair, Dhanalakshmi R, Occhipinti, Patricia, Borsuk, Mark E, Gladfelter, Amy S, Pathogen, The Plant, Roberts, Samantha E, Gladfelter, Amy S, Roper, Marcus, Lee, Changhwan, Hickey, Patrick C, Gladfelter, Amy S, Jose, Juan, Jose, Francisco, Tibayrenc, Michel, & Diosque, Patricio. 2013. Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. *Current Opinion in Microbiology*, **10**(5), 12875–80.

Gabaldón, Toni, Naranjo-Ortíz, Miguel A., & Marcet-Houben, Marina. 2016. Evolutionary genomics of yeast pathogens in the Saccharomycotina. *FEMS yeast research*, **16**(6), fow064.

Galeote, V., Novo, M., Salema-Oom, M., Brion, C., Valerio, E., Goncalves, P., & Dequin, S. 2010. FSY1, a horizontally transferred gene in the Saccharomyces cerevisiae EC1118 wine yeast strain, encodes a high-affinity fructose/H+ symporter. *Microbiology*, **156**(12), 3754–3761.

Gallone, Brigida, Steensels, Jan, Prahl, Troels, Soriaga, Leah, Saels, Veerle, Herrera-Malaver, Beatriz, Merlevede, Adriaan, Roncoroni, Miguel, Voordeckers, Karin, Miraglia, Loren, Teiling, Clotilde, Steffy, Brian, Taylor, Maryann, Schwartz, Ariel, Richardson, Toby, White, Christopher, Baele, Guy, Maere, Steven, & Verstrepen, Kevin J. 2016. Domestication and Divergence of Saccharomyces cerevisiae Beer Yeasts. *Cell*, **166**(6), 1397–1410.e16.

Galtier, Nicolas, & Daubin, Vincent. 2008. Dealing with incongruence in phylogenomic analyses. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **363**(1512), 4023–4029.

Garbelotto, Matteo, & Gonthier, Paolo. 2013. Biology, Epidemiology, and Control of Heterobasidion Species Worldwide. *Annual Review of Phytopathology*, **51**(1), 39–59.

Garbelotto, Matteo, Gonthier, Paolo, Linzer, Rachel, Nicolotti, Giovanni, & Otrosina, William. 2004. A shift in nuclear state as the result of natural interspecific hybridization between two North American taxa of the basidiomycete complex Heterobasidion. *Fungal genetics and biology : FG & B*, **41**(11), 1046–51.

Garcia-Vallvé, S., Romeu, A., & Palau, J. 2000. Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. *Molecular biology and evolution*, **17**(3), 352–61.

Gerstein, Aleeza C, & Berman, Judith. 2015. Shift and adapt: the costs and benefits of karyotype variations. *Current opinion in microbiology*, **26**(aug), 130–6.

Gerstein, Aleeza C, Fu, Man Shun, Mukaremera, Liliane, Li, Zhongming, Ormerod, Kate L, Fraser, James A, Berman, Judith, & Nielsen, Kirsten. 2015. Polyploid titan cells produce haploid and aneuploid progeny to promote stress adaptation. *mBio*, **6**(5), e01340–15.

Gibeaux, Romain, Politi, Antonio Z, Philippsen, Peter, & Nédélec, François. 2017. Mechanism of nuclear movements in a multinucleated cell. *Molecular biology of the cell*, **28**(5), 645–660.

Gilbert, Matthew K., Mack, Brian M., Moore, Geromy G., Downey, Darlene L., Lebar, Matthew D., Joardar, Vinita, Losada, Liliana, Yu, JiuJiang, Nierman, William C., & Bhatnagar, Deepak. 2018. Whole genome comparison of Aspergillus flavus L-morphotype strain NRRL 3357 (type) and S-morphotype strain AF70. *PLOS ONE*, **13**(7), e0199169.

Giordano, Luana, Sillo, Fabiano, Garbelotto, Matteo, & Gonthier, Paolo. 2018. Mitonuclear interactions may contribute to fitness of fungal hybrids. *Scientific Reports*, **8**(1706).

Gladfelter, Amy S. 2006. Nuclear anarchy: asynchronous mitosis in multinucleated fungal hyphae. *Current Opinion in Microbiology*, **9**(6), 547–552.

Gluck-Thaler, Emile, Slot, Jason C., Thomas, CM, Nielsen, KM, Baquero, F, Smillie, CS, Smith, MB, Friedman, J, Cordero, OX, David, LA, Barlow, M, Kroken, S, Glass, NL, Taylor, JW, Yoder, OC, Turgeon, BG, Soanes, D, Richards, TA, Zhaxybayeva, O, Doolittle, WF, Heath, TA, Hedtke, SM, Hillis, DM, James, TY, Pelin, A, Bonen, L, Ahrendt, S, Sain, D, Lin, K, Limpens, E, Zhang, Z, Ivanov, S, Saunders, DGO, Sun, G, Yang, Z, Kosch, T, Summers, K, Huang, J, Campbell, MA, Staats, M, van Kan, JA, Rokas, A, Slot, JC, Grigoriev, IV, Nikitin, R, Haridas, S, Kuo, A, Ohm, R, Jain, R, Rivera, MC, Lake, JA, Pal, C, Papp, B, Lercher, MJ, Cohen, O, Gophna, U, Pupko, T, Friesen, TL, Stukenbrock, EH, Liu, Z, Meinhardt, S, Ling, H, Richards, TA, Soanes, DM, Jones, MDM, Vasieva, O, Leonard, G, Klosterman, SJ, Subbarao, KV, Kang, S, Veronese, P, Gold, SE, Tsaousis, AD, de Choudens, S Ollagnier, Gentekaki, E, Long, S, Gaston, D, Greene, GH, McGary, KL, Rokas, A, Slot, JC, Khaldi, N, Wolfe, KH, Patron, NJ, Waller, RF, Cozijnsen, AJ, Straney, DC, Gardiner, DM, Elmore, MH, McGary, KL, Wisecaver, JH, Slot, JC, Geiser, DM, Carlton, JM, Hirt, RP, Silva, JC, Delcher, AL, Schatz, M, Strese, A, Backlund, A, Alsmark, C, Whitaker, JW, McConkey, GA, Westhead, DR, Wisecaver, JH, Slot, JC, Rokas, A, Sun, B-F, Xiao, J-H, He, S, Liu, L, Murphy, RW, Richards, TA, Talbot, NJ, Savory, F, Leonard, G, Richards, TA, Zhao, H, Xu, C, Lu, H-L, Chen, X, Leger, RJ St., Pombert, J-F, Haag, KL, Beidas, S, Ebert, D, Keeling, PJ,

Gardiner, DM, McDonald, MC, Covarelli, L, Solomon, PS, Rusu, AG, Slot, JC, Hibbett, DS, Slot, JC, Rokas, A, Selman, M, Pombert, JF, Solter, L, Farinelli, L, Weiss, LM, Jaramillo, VD Armijos, Vargas, WA, Sukno, SA, Thon, MR, Doolittle, W Ford, Walther, BA, Ewald, PW, Parmesan, C, & Yohe, G. 2015. Dimensions of Horizontal Gene Transfer in Eukaryotic Microbial Pathogens. *PLOS Pathogens*, **11**(10), e1005156.

Gojkovic, Z, Knecht, W, Zameitat, E, Warneboldt, J, Coutelis, J.-B, Pynyaha, Y, Neuveglise, C, Moller, K, Loffer, M, & Piskur, J. 2004. Horizontal gene transfer promoted evolution of the ability to propagate under anaerobic conditions in yeasts. *Molecular Genetics and Genomics*, **271**, 387–393.

Golicz, Agnieszka A, Bayer, Philipp E, Barker, Guy C, Edger, Patrick P, Kim, HyeRan, Martinez, Paula A, Chan, Chon Kit Kenneth, Severn-Ellis, Anita, McCombie, W Richard, Parkin, Isobel A P, Paterson, Andrew H, Pires, J Chris, Sharpe, Andrew G, Tang, Haibao, Teakle, Graham R, Town, Christopher D, Batley, Jacqueline, & Edwards, David. 2016a. The pangenome of an agronomically important crop plant Brassica oleracea. *Nature communications*, **7**(nov), 13390.

Golicz, Agnieszka A., Batley, Jacqueline, & Edwards, David. 2016b. Towards plant pangenomics. *Plant Biotechnology Journal*, **14**(4), 1099–1105.

Gonthier, Paolo, Warner, Rachel, Nicolotti, Giovanni, Mazzaglia, Angelo, & Garbelotto, Matteo M. 2004. Pathogen Introduction as A Collateral Effect of Military Activity. *Mycological Research*, **108**(5), 468–470.

Grant, J. R., & Katz, L. a. 2014. Phylogenomic study indicates widespread lateral gene transfer in Entamoeba and suggests a past intimate relationship with parabasalids. *Genome Biology and Evolution*, **6**(9), 2350–2360.

Haegeman, Annelies, Jones, John T, Danchin, Etienne G J, Hall, Charles, Brachat, Sophie, Dietrich, Fred S, Huang, Jinling, Mullapudi, Nandita, Lancto, Cheryl A, Scott, Marla, Abrahamsen, Mitchell S, Kissinger, Jessica C, Schönknecht, Gerald, Weber, Andreas P M, Lercher, Martin J, Domazet-Lošo, Mirjana, Haubold, Bernhard, Gladyshev, Eugene a, Meselson, Matthew, Arkhipova, Irina R, Slot, Jason C, Hibbett, David S, Andersson, Jan O., Hirt, Robert P, Foster, Peter G, Roger, Andrew J, Bruto, Maxime, Prigent-combaret, Claire, Luis, Patricia, Muller, Daniel, Chan, Cheong Xin, Bhattacharya, Debashish, Leigh, Jessica W, Lapointe, François-Joseph, Lopez, Philippe, Bapteste, Eric, Richards, Thomas A., Soanes, Darren M., Foster, Peter G, Leonard, Guy, Thornton, Christopher R, Talbot, Nicholas J., Fitzpatrick, David a, Choi, In-Geol, Kim, Sung-Hou, Boto, Luis, Dunning Hotopp, Julie C, Emiliani, Giovanni, Fondi, Marco, Fani, Renato, Gribaldo, Simonetta, Richards, Thomas A., Dacks, Joel B, Jenkinson, Joanna M, Thornton, Christopher R, Talbot, Nicholas J., Bloomquist, Erik W, Suchard, Marc a, Bock, Ralph, Moriguchi, Kazuki, Yamamoto, Shinji, Tanaka, Katsuyuki, Kurata, Nori, Suzuki, Katsunori, Gao, Caihua, Ren, Xiaodong, Mason, Annaliese S, Liu, Honglei, Xiao, Meili, Li, Jiana, Fu, Donghui, Garcia-Vallvé, S, Romeu, a, Palau, J, Grant, J. R., Katz, Laura a., Keeling, Patrick J, Sun, Bao-Fa, Xiao, Jin-Hua, He, Shunmin, Liu, Li, Murphy, Robert W, Huang, Da-Wei, Hall, Charles, Dietrich, Fred S, Marcet-Houben, Marina, Gabaldón, Toni, Whitaker, John W, McConkey, Glenn a, Westhead, David R, Schulte, R D, Makus, C, Schulenburg, H, Jackson, Daniel J, Macis, Luciana, Reitner, Joachim, Wörheide, Gert, Wisecaver, Jennifer H, Brosnahan, Michael L, Hackett, Jeremiah D, Keeling, Patrick J, Palmer, Jeffrey D, Khaldi, Nora, Collemare, Jérôme, Lebrun, Marc-Henri, Wolfe, Kenneth H, Moran, Nancy a, Jarvik, Tyler, Fitzpatrick, David a, Logue, Mary E, Butler, Geraldine, Andersson, Jan O., Richards, Thomas A., Leonard, Guy, Soanes, Darren M., Talbot, Nicholas J., Qiu, Huan, Yoon, Hwan Su, Bhattacharya, Debashish, Nikolaidis, Nikolas, Doran, Nicole, Cosgrove, Daniel J, Oliveira, Sarah G, Bao, Weidong, Martins, Cesar, Jurka, Jerzy, Richards, Thomas A., Soanes, Darren M., Jones, Meredith D M, Vasieva, Olga, Leonard, Guy, Talbot, Nicholas J., Richardson, Aaron O, Palmer, Jeffrey D, Rosewich, U Liane, Kistler, H Corby, Rybarczyk-Mydłowska, Katarzyna, Maboreke, Hazel Ruvimbo, van Megen, Hanny, van den Elsen, Sven, Mooyman, Paul, Smant, Geert, Bakker, Jaap, Helder, Johannes, Sanders, Ian R, Schmitt, Imke, Lumbsch, H Thorsten, Schönknecht, Gerald, Chen, Wei-Hua, Ternes, Chad M, Barbier, Guillaume G, Shrestha, Roshan P, Stanke, Mario, Bräutigam, Andrea, Baker, Brett J, Banfield, Jillian F, Garavito, R Michael, Carr, Kevin, Wilkerson, Curtis, Rensing, Stefan a, Gagneul, David, Dickenson, Nicholas E, Oesterhelt, Christine, Lercher, Martin J, Weber, Andreas P M, Slot, Jason C, Rokas, Antonis, Soanes, Darren M., Richards, Thomas A., Strese, Ake, Backlund, Anders, Alsmark, Cecilia, Taylor, Mark, Mediannikov, Oleg, Raoult, Didier, Greub, Gilbert, Ternes, Chad M, Schönknecht, Gerald, Wallau, Gabriel Luz, Ortiz, Mauro Freitas, Loreto, Elgion Lucio Silva, Wijayawardena, Bhagya K, Minchella, Dennis J, Dewoody, J Andrew, Won, Hyosig, Renner, Susanne S, Doolittle, W. Ford, Tucker, Richard P, Denker, Elsa, Bapteste, Eric, Guyader, Hervé Le, Manuel, Michaël, Rabet, Nicolas, Whitaker, John W, McConkey, Glenn a, Westhead, David R, Walther, Andrea, Hesselbart, Ana, Ambrose, Karen V, Koppenho, Albrecht M, Belanger, Faith C, Baltrus, David A, Betat, Heike, Mede, Tobias, Tretbar, Sandy, Steiner, Lydia, Stadler, Peter F, Prohaska, Sonja J, Boothby, Thomas C, Tenlen, Jennifer R, Smith, Frank W, Wang, Jeremy R, Patanella, Kiera A, Osborne, Erin, Tintori, Sophia C, Li, Qing, Jones, Corbin D, Yandell, Mark, Messina, David N, Glasscock, Jarret, Goldstein, Bob, Boschetti, Chiara, Carr, Adrian, Crisp, Alastair, Eyres, Isobel, Wang-koh, Yuan, Lubzens, Esther, Barraclough, Timothy G, Micklem, Gos, Tunnacliffe, Alan, Cohen, Ofir, Pupko, Tal, Crisp, Alastair, Dana, Catherine E, Glauber, Kristine M, Chan, Titus A, Bridge, Diane M, Steele, Robert E, Davis, Charles C, Xi, Zhenxiang, Gonc, Paula, Coelho, Marco A, Gonc, Carla, Remarks, Concluding, Republic, Czech, Republic, Czech, Wingfield, Brenda Diana, Baidouri, Moaine El, Carpentier, Marie-christine, Cooke, Richard, Gao, Dongying, Lasserre, Eric, Llauro, Christel, Mirouze, Marie, Picault, Nathalie, Jackson, Scott A, Panaud, Olivier, Gojkovic, Z,

Lo, M, Transfer, Horizontal, Hespeels, Boris, Li, Xueqiong Xiang, Flot, Jean-françois, Pigneur, Lise-marie, Malaisse, Jeremy, Silva, Corinne Da, Doninck, Karine Van, Access, Open, Danilo, Vinicio, Jaramillo, Armijos, Vargas, Walter Alberto, Sukno, Serenella Ana, Thon, Michael R, Hirt, Robert P, Alsmark, Cecilia, Embley, T Martin, Ivancevic, Atma M, Walsh, Ali M, Kortschak, R Daniel, Adelson, David L, Katz, Laura a., Kirsch, Roy, Gramzow, Lydia, Theissen, Günter, Siegfried, Blair D, Richard, H, Heckel, David G, Pauchet, Yannick, Kyndt, Tina, Quispe, Dora, Zhai, Hong, Jarret, Robert, Ghislain, Marc, Liu, Qingchang, Gheysen, Godelieve, Kreuze, Jan F, List, Johann-mattis, Nelson-sathi, Shijulal, Geisler, Hans, Martin, William, Martijn, Joran, Ettema, Thijs J G, Modolo, Laurent, Picard, Franck, Lerat, Emmanuelle, Belkorchia, Abdel, Panek, Johan, Parisot, Nicolas, Pelin, Adrian, Gasc, Cyrielle, Alaoui, Hicham El, Biron, David G, Vaury, Chantal, Peyret, Pierre, Corradi, Nicolas, Lerat, Emmanuelle, Nguyen, Marcus, Ekstrom, Alex, Li, Xueqiong Xiang, Yin, Yanbin, Rödelsperger, Christian, Sommer, Ralf J, Brief, In, Numbers, Accession, Shterzer, N, Mizrahi, I, Soucy, Shannon M, Huang, Jinling, Gogarten, Johann Peter, Stiller, John W, Suzuki, Katsunori, Moriguchi, Kazuki, Yamamoto, Shinji, Syvanen, Michael, Szydlowski, L, Boschetti, Chiara, Crisp, Alastair, Barbosa, E G G, Tunnacliffe, Alan, Whitehead, Michael P, Hooley, Paul, Brown, Michael R W, Wijayawardena, Bhagya K, Minchella, Dennis J, Dewoody, J Andrew, Wisecaver, Jennifer H, Hackett, Jeremiah D, Zhang, Miaozhi, Maryam, Chaib De Mares, & Elsas, Jan Dirk Van. 2014. Horizontal gene transfer in plants. *BMC evolutionary biology*, **4**(1), 1–8.

Hallatschek, Oskar, & Nelson, David R. 2008. Gene surfing in expanding populations. *Theoretical Population Biology*, **73**(1), 158–170.

Hamilton, Cyd E, Faeth, Stan H, & Dowling, Thomas E. 2009. Distribution of hybrid fungal symbionts and environmental stress. *Microbial ecology*, **58**(2), 408–13.

Hane, James K, Rouxel, Thierry, Howlett, Barbara J, Kema, Gert Hj, Goodwin, Stephen B, & Oliver, Richard P. 2011. A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biology*, **12**, R45.

Harrison, Benjamin D, Hashemi, Jordan, Bibi, Maayan, Pulver, Rebecca, Bavli, Danny, Nahmias, Yaakov, Wellington, Melanie, Sapiro, Guillermo, & Berman, Judith. 2014. A tetraploid intermediate precedes aneuploid formation in yeasts exposed to fluconazole. *PLoS biology*, **12**(3), e1001815.

Heinemann, Jack A., & Sprague, George F. 1989. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature*, **340**(6230), 205–209.

Heinemann, Jack A., & Sprague, George F. 1991. Transmission of Plasmid DNA to Yeast by Conjugation with Bacteria. *Methods in Enzymology*, **194**(C), 187–195.

Hellborg, Linda, & Piskur, Jure. 2009. Complex nature of the genome in a wine spoilage yeast, Dekkera bruxellensis. *Eukaryotic cell*, **8**(11), 1739–49.

Hickman, Meleah A, Paulson, Carsten, Dudley, Aimee, & Berman, Judith. 2015. Parasexual Ploidy Reduction Drives Population Heterogeneity Through Random and Transient Aneuploidy in Candida albicans. *Genetics*, **200**(3), 781–94.

Hittinger, Chris Todd. 2013. Saccharomyces diversity and evolution: a budding model genus. *Trends in Genetics*, **29**(5), 309–317.

Huang, Shengfeng, Kang, Mingjing, & Xu, Anlong. 2017. HaploMerger2: rebuilding both haploid subassemblies from high-heterozygosity diploid genome assembly. *Bioinformatics*, **490**(apr), 49–54.

Hur, Gene H, Vickery, Christopher R, & Burkart, Michael D. 2012. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Natural product reports*, **29**(10), 1074–98.

Ibarra, Rafael U., Edwards, Jeremy S., & Palsson, Bernhard O. 2002. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, **420**(6912), 186–189.

Inomata, K., Nishikawa, M., & Yoshida, K. 1994. *The yeast Saccharomyces kluyveri as a recipient eukaryote in transkingdom conjugation: Behavior of transmitted plasmids in transconjugants.*

Ishikawa, Francine H, Souza, Elaine A, Shoji, Jun-Ya, Connolly, Lanelle, Freitag, Michael, Read, Nick D, & Roca, M Gabriela. 2012. Heterokaryon incompatibility is suppressed following conidial anastomosis tube fusion in a fungal plant pathogen. *PloS one*, **7**(2), e31175.

Iwasa, Tanabe, & Kamada. 1998. The Two Nuclei in the Dikaryon of the Homobasidiomycete Coprinus cinereus Change Position after Each Conjugate Division. *Fungal genetics and biology*, **23**(1), 110–6.

Jain, R., Rivera, M. C., & Lake, J. A. 1999. Horizontal gene transfer among genomes: The complexity hypothesis. *Proceedings of the National Academy of Sciences*, **96**(7), 3801–3806.

James, Stephen a., Bond, Christopher J., Stratford, Malcolm, & Roberts, Ian N. 2005. Molecular evidence for the existence of natural hybrids in the genus Zygosaccharomyces. *FEMS Yeast Research*, **5**(8), 747–755.

James, Timothy Y., Stenlid, Jan, Olson, Åke, & Johannesson, Hanna. 2008. Evolutionary significance of imbalanced nuclear ratios within heterokaryons of the basidiomycete fungus Heterobasidion parviporum. *Evolution*, **62**(9), 2279–2296.

James, Timothy Y., Johansson, Stina B.K., & Johannesson, Hanna. 2009. Trikaryon formation and nuclear selection in pairings between heterokaryons and homokaryons of the root rot pathogen Heterobasidion parviporum. *Mycological Research*, **113**(5), 583–590.

Jang, J, Hur, H.-G, Sadowsky, M J, Byappanahalli, M N, Yan, T, & Ishii, S. 2017. Environmental Escherichia coli: ecology and public health implications-a review. *Journal of Applied Microbiology*, **123**, 570–581.

Jany, Jean-Luc, & Pawlowska, Teresa E. E. 2010. Multinucleate spores contribute to evolutionary longevity of asexual glomeromycota. *The American naturalist*, **175**(4), 424–35.

Jiang, Dewei, Zhu, Wei, Wang, Yunchuan, Sun, Chang, Zhang, Ke-Qin, & Yang, Jinkui. 2013. Molecular tools for functional genomics in filamentous fungi: Recent advances and new strategies. *Biotechnology Advances*, **31**(8), 1562–1574.

Jiao, Yuannian, Wickett, Norman J, Ayyampalayam, Saravanaraj, Chanderbali, André S, Landherr, Lena, Ralph, Paula E, Tomsho, Lynn P, Hu, Yi, Liang, Haiying, Soltis, Pamela S, Soltis, Douglas E, Clifton, Sandra W, Schlarbaum, Scott E, Schuster, Stephan C, Ma, Hong, Leebens-Mack, Jim, & DePamphilis, Claude W. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature*, **473**(7345), 97–100.

Jinks, J. L. 1952. Heterocaryosis in wild penicillium. *Heredity*, **6**(1), 77–87.

Johannesson, Hanna, & Samils, Nicklas. 2014. Nuclear interactions in a heterokaryon: insight from the model Neurospora tetrasperma. *Proceedings of the Royal Society Biological sciences*, **281**(20140084).

Joneson, Suzanne, Stajich, Jason E, Shiu, Shin-Han, & Rosenblum, Erica Bree. 2011. Genomic transition to pathogenicity in chytrid fungi. *PLoS pathogens*, **7**(11), e1002338.

Kajitani, Rei, Toshimoto, Kouta, Noguchi, Hideki, Toyoda, Atsushi, Ogura, Yoshitoshi, Okuno, Miki, Yabana, Mitsuru, Harada, Masayuki, Nagayasu, Eiji, Maruyama, Haruhiko, Kohara, Yuji, Fujiyama, Asao, Hayashi, Tetsuya, & Itoh, Takehiko. 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Research*, **24**(8), 1384–1395.

Katz, Laura A. 2015. Recent events dominate interdomain lateral gene transfers between prokaryotes and eukaryotes and, with the exception of endosymbiotic gene transfers, few ancient transfer events persist. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **370**(1678), 20140324.

Kavanagh, Kevin, & Whittaker, Peter A. 1996. Application of protoplast fusion to the nonconventional yeast. *Enzyme and Microbial Technology*, **18**(1), 45–51.

Kelly, Amy C., & Ward, Todd J. 2018. Population genomics of Fusarium graminearum reveals signatures of divergent evolution within a major cereal pathogen. *PLOS ONE*, **13**(3), e0194616.

Kenny, N J, Chan, K W, Nong, W, Qu, Z, Maeso, I, Yip, H Y, Chan, T F, Kwan, H S, Holland, P W H, Chu, K H, & Hui, J H L. 2016. Ancestral whole-genome duplication in the marine chelicerate horseshoe crabs. *Heredity*, **116**(2), 190–9.

Khaldi, Nora, Collemare, Jérôme, Lebrun, Marc-Henri, & Wolfe, Kenneth H. 2008. Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. *Genome biology*, **9**(1), R18.

Kinsey, J A. 1990. Tad, a LINE-like transposable element of Neurospora, can transpose between nuclei in heterokaryons. *Genetics*, **126**(2), 317–23.

Kravets, Anatoliy, Yang, Feng, Bethlendy, Gabor, Sherman, Fred, & Rustchenko, Elena. 2014. Adaptation of Candida albicans to growth on sorbose via monosomy of chromosome 5 accompanied by duplication of another chromosome carrying a gene responsible for sorbose utilization. *FEMS Yeast Research*, **14**(5), 708–713.

Krishnan, Parvathy, Ma, Xin, McDonald, Bruce A, & Brunner, Patrick C. 2018. Widespread signatures of selection for secreted peptidases in a fungal plant pathogen. *BMC evolutionary biology*, **18**(1), 7.

Kumaran, Rajaraman, Yang, Shi Yow, & Leu, Jun Yi. 2013. Characterization of Chromosome Stability in Diploid, Polyploid and Hybrid Yeast Cells. *PLoS ONE*, **8**(7).

Kuo, Alan, Kohler, Annegret, Martin, Francis M, & Grigoriev, Igor V. 2014. Expanding genomics of mycorrhizal symbiosis. *Frontiers in microbiology*, **5**, 582.

Lacroix, Benoît, & Citovsky, Vitaly. 2016. Transfer of DNA from Bacteria to Eukaryotes. *mBio*, **7**(4), e00863–16.

Lacroix, Benoît, Tzfira, Tzvi, Vainstein, Alexander, & Citovsky, Vitaly. 2006. *A case of promiscuity: Agrobacterium's endless hunt for new partners*.

Lander, E S, Linton, L M, Birren, B, Nusbaum, C, Zody, M C, Baldwin, J, Devon, K, Dewar, K, Doyle, M, FitzHugh, W, Funke, R, Gage, D, Harris, K, Heaford, A, Howland, J, Kann, L, Lehoczky, J, LeVine, R, McEwan, P, McKernan, K, Meldrim, J, Mesirov, J P, Miranda, C, Morris, W, Naylor, J, Raymond, C, Rosetti, M, Santos, R, Sheridan, A, Sougnez, C, Stange-Thomann, N, Stojanovic, N, Subramanian, A, Wyman, D, Rogers, J, Sulston, J, Ainscough, R, Beck, S, Bentley, D, Burton, J, Clee, C, Carter, N, Coulson, A, Deadman, R, Deloukas, P, Dunham, A, Dunham, I, Durbin, R, French, L, Grafham, D, Gregory, S, Hubbard, T, Humphray, S, Hunt, A, Jones, M, Lloyd, C, McMurray, A, Matthews, L, Mercer, S, Milne, S, Mullikin, J C, Mungall, A, Plumb, R, Ross, M, Shownkeen, R, Sims, S, Waterston, R H, Wilson, R K, Hillier, L W, McPherson, J D, Marra, M A, Mardis, E R, Fulton, L A, Chinwalla, A T, Pepin, K H, Gish, W R, Chissoe, S L, Wendl, M C, Delehaunty, K D, Miner, T L, Delehaunty, A, Kramer, J B, Cook, L L, Fulton, R S, Johnson, D L, Minx, P J, Clifton, S W, Hawkins, T, Branscomb, E, Predki, P, Richardson, P, Wenning, S, Slezak, T, Doggett, N, Cheng, J F, Olsen, A, Lucas, S, Elkin, C, Uberbacher, E, Frazier, M, Gibbs, R A, Muzny, D M, Scherer, S E, Bouck, J B, Sodergren, E J, Worley, K C, Rives, C M, Gorrell, J H, Metzker, M L, Naylor, S L, Kucherlapati, R S, Nelson, D L, Weinstock, G M, Sakaki, Y, Fujiyama, A, Hattori, M, Yada, T, Toyoda, A, Itoh, T, Kawagoe, C, Watanabe, H, Totoki, Y, Taylor, T, Weissenbach, J, Heilig, R, Saurin, W, Artiguenave, F, Brottier, P, Bruls, T, Pelletier, E, Robert, C, Wincker, P, Smith, D R, Doucette-Stamm, L, Rubenfield, M, Weinstock, K, Lee, H M, Dubois, J, Rosenthal, A, Platzer, M, Nyakatura, G, Taudien, S, Rump, A, Yang, H, Yu, J, Wang, J, Huang, G, Gu, J, Hood, L, Rowen, L, Madan, A, Qin, S, Davis, R W, Federspiel, N A, Abola, A P, Proctor, M J, Myers, R M, Schmutz, J, Dickson, M, Grimwood, J, Cox, D R, Olson, M V, Kaul, R, Raymond, C, Shimizu, N, Kawasaki, K, Minoshima, S, Evans, G A, Athanasiou, M, Schultz, R, Roe, B A, Chen, F, Pan, H, Ramser, J, Lehrach, H, Reinhardt, R, McCombie, W R, de la Bastide, M, Dedhia, N, Blocker, H, Hornischer, K, Nordsiek, G, Agarwala, R, Aravind, L, Bailey, J A, Bateman, A, Batzoglou, S, Birney, E, Bork, P, Brown, D G, Burge, C B, Cerutti, L, Chen, H C, Church, D, Clamp, M, Copley, R R, Doerks, T, Eddy, S R, Eichler, E E, Furey, T S, Galagan, J, Gilbert, J G, Harmon, C, Hayashizaki, Y, Haussler, D, Hermjakob, H, Hokamp, K, Jang, W, Johnson, L S, Jones, T A, Kasif, S, Kaspryzk, A, Kennedy, S, Kent, W J, Kitts, P,

Koonin, E V, Korf, I, Kulp, D, Lancet, D, Lowe, T M, McLysaght, A, Mikkelsen, T, Moran, J V, Mulder, N, Pollara, V J, Ponting, C P, Schuler, G, Schultz, J, Slater, G, Smit, A F, Stupka, E, Szustakowski, J, Thierry-Mieg, D, Thierry-Mieg, J, Wagner, L, Wallis, J, Wheeler, R, Williams, A, Wolf, Y I, Wolfe, K H, Yang, S P, Yeh, R F, Collins, F, Guyer, M S, Peterson, J, Felsenfeld, A, Wetterstrand, K A, Patrinos, A, Morgan, M J, de Jong, P, Catanese, J J, Osoegawa, K, Shizuya, H, Choi, S, Chen, Y J, & International Human Genome Sequencing, Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature*, **409**(6822), 860–921.

Leducq, Jean-Baptiste, Nielly-Thibault, Lou, Charron, Guillaume, Eberlein, Chris, Verta, Jukka-Pekka, Samani, Pedram, Sylvester, Kayla, Hittinger, Chris Todd, Bell, Graham, & Landry, Christian R. 2016. Speciation driven by hybridization and chromosomal plasticity in a wild yeast. *Nature microbiology*, **1**(jan), 15003.

Leigh, Jessica W, Lapointe, François-Joseph, Lopez, Philippe, & Bapteste, Eric. 2011. Evaluating phylogenetic congruence in the post-genomic era. *Genome biology and evolution*, **3**(jan), 571–87.

Leitch, A R, & Leitch, I J. 2012. Ecological and genetic factors linked to contrasting genome dynamics in seed plants. *New phytologist*, **194**, 629–646.

Lenassi, Metka, Gostinčar, Cene, Jackman, Shaun, Turk, Martina, Sadowski, Ivan, Nislow, Corey, Jones, Steven, Birol, Inanc, Cimerman, Nina Gunde, & Plemenitaš, Ana. 2013. Whole Genome Duplication and Enrichment of Metal Cation Transporters Revealed by De Novo Genome Sequencing of Extremely Halotolerant Black Yeast Hortaea werneckii. *PLoS ONE*, **8**(8), e71328.

Li, Wenjun, Averette, Anna Floyd, Desnos-Ollivier, Marie, Ni, Min, Dromer, Françoise, & Heitman, Joseph. 2012. Genetic Diversity and Genomic Plasticity of Cryptococcus neoformans AD Hybrid Strains. *G3: Genes, Genomes, Genetics*, **2**(1), 83–97.

Li, Zheng, Tiley, George P., Galuska, Sally R., Reardon, Chris R., Kidder, Thomas I., Rundell, Rebecca J., & Barker, Michael S. 2018b. Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proceedings of the National Academy of Sciences*, **115**(18), 201710791.

Lin, Kui, Limpens, Erik, Zhang, Zhonghua, Ivanov, Sergey, Saunders, Diane G. O., Mu, Desheng, Pang, Erli, Cao, Huifen, Cha, Hwangho, Lin, Tao, Zhou, Qian, Shang, Yi, Li, Ying, Sharma, Trupti, van Velzen, Robin, de Ruijter, Norbert, Aanen, Duur K., Win, Joe, Kamoun, Sophien, Bisseling, Ton, Geurts, René, & Huang, Sanwen. 2014a. Single Nucleus Genome Sequencing Reveals High Similarity among Nuclei of an Endomycorrhizal Fungus. *PLoS Genetics*, **10**(1), e1004078.

Ma, Li-Jun, Ibrahim, Ashraf S, Skory, Christopher, Grabherr, Manfred G, Burger, Gertraud, Butler, Margi, Elias, Marek, Idnurm, Alexander, Lang, B Franz, Sone, Teruo, Abe, Ayumi, Calvo, Sarah E, Corrochano, Luis M, Engels, Reinhard, Fu, Jianmin, Hansberg, Wilhelm, Kim, Jung-Mi, Kodira, Chinnappa D, Koehrsen, Michael J, Liu, Bo, Miranda-Saavedra, Diego, O'Leary, Sinead, Ortiz-Castellanos, Lucila, Poulter, Russell, Rodriguez-Romero, Julio, Ruiz-Herrera, José, Shen, Yao-Qing, Zeng, Qiandong, Galagan, James, Birren, Bruce W, Cuomo, Christina A, & Wickes, Brian L. 2009. Genomic analysis of the basal lineage fungus Rhizopus oryzae reveals a whole-genome duplication. *PLoS genetics*, **5**(7), e1000549.

Magadum, Santoshkumar, Banerjee, Urbi, Murugan, Priyadharshini, Gangapur, Doddabhimappa, & Ravikesavan, Rajasekar. 2013. Gene duplication as a major force in evolution. *J. Genet*, **92**, 155–161.

Maheshwari, Ramesh. 2005. Nuclear behavior in fungal hyphae. *FEMS Microbiology Letters*, **249**, 7–14.

Manning, Robert J., & Callaghan, Arthur A. 2008. Pathogenicity of Conidiobolus spp. and Basidiobolus ranarum to arthropods co-occurring in leaf litter. *Fungal Ecology*, **1**(1), 33–39.

Manning, Robert J., Waters, Steve D., & Callaghan, Arthur A. 2007. Saprotrophy of Conidiobolus and Basidiobolus in leaf litter. *Mycological Research*, **111**(12), 1437–1449.

Marahiel, Mohamed a. 2009. Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *Journal of peptide science*, **15**(12), 799–807.

Marcet-Houben, Marina, & Gabaldón, Toni. 2010. Acquisition of prokaryotic genes by fungal genomes. *Trends in genetics : TIG*, **26**(1), 5–8.

Marcet-houben, Marina, & Gabaldón, Toni. 2015. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. *PLoS biology*, **13**(8), 1–26.

Marsit, Souhir, Mena, Adriana, Bigey, Frédéric, Sauvage, François-Xavier, Couloux, Arnaud, Guy, Julie, Legras, Jean-Luc, Barrio, Eladio, Dequin, Sylvie, & Galeote, Virginie. 2015. Evolutionary Advantage Conferred by an Eukaryote-to-Eukaryote Gene Transfer Event in Wine Yeasts. *Molecular biology and evolution*, **32**(7), 1695–707.

Mathieu, Stephanie, Cusant, Loïc, Roux, Christophe, & Corradi, Nicolas. 2018. Arbuscular mycorrhizal fungi: intraspecific diversity and pangenomes. *New Phytologist*, jun.

Mixão, Verónica, & Gabaldón, Toni. 2018. Yeast Interspecies Hybrids Hybridization and emergence of virulence in opportunistic human yeast pathogens. *Yeast*, **35**, 5–20.

Möller, Mareike, & Stukenbrock, Eva H. 2017. Evolution and genome architecture in fungal plant pathogens. *Nature Reviews Microbiology*, **15**(12), 756–771.

Morales, Lucia, & Dujon, Bernard. 2012. Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. *Microbiology and molecular biology reviews : MMBR*, **76**(4), 721–39.

Moriguchi, Kazuki, Yamamoto, Shinji, Tanaka, Katsuyuki, Kurata, Nori, & Suzuki, Katsunori. 2013. Trans-kingdom horizontal DNA transfer from bacteria to yeast is highly plastic due to natural polymorphisms in auxiliary nonessential recipient genes. *PloS one*, **8**(9), e74590.

Morris, J. Jeffrey. 2015. *Black Queen evolution: The role of leakiness in structuring microbial communities.*

Morrow, Carl a., & Fraser, James a. 2013. Ploidy variation as an adaptive mechanism in human pathogenic fungi. *Seminars in Cell and Developmental Biology*, **24**(4), 339–346.

Nair, Dhanalakshmi R, Occhipinti, Patricia, Borsuk, Mark E, & Gladfelter, Amy S. 2010. A conserved

G 1 regulatory circuit promotes asynchronous behavior of nuclei sharing a common cytoplasm. *Cell Cycle*, **9**(18), 3771–3779.

Naranjo-Ortíz, Miguel A., Brock, Matthias, Brunke, Sascha, Hube, Bernhard, Marcet-Houben, Marina, & Gabaldón, Toni. 2016. Widespread Inter- and Intra-Domain Horizontal Gene Transfer of d-Amino Acid Metabolism Enzymes in Eukaryotes. *Frontiers in Microbiology*, **7**(dec), 2001.

Neafsey, Daniel E, Barker, Bridget M, Sharpton, Thomas J, Stajich, Jason E, Park, Daniel J, Whiston, Emily, Hung, Chiung-Yu, Mcmahan, Cody, White, Jared, Sykes, Sean, Heiman, David, Young, Sarah, Zeng, Qiandong, Abouelleil, Amr, Aftuck, Lynne, Bessette, Daniel, Brown, Adam, FitzGerald, Michael, Lui, Annie, Macdonald, J Pendexter, Priest, Margaret, Orbach, Marc J, Galgiani, John N, Kirkland, Theo N, Cole, Garry T, Birren, Bruce W, Henn, Matthew R, Taylor, John W, & Rounsley, Steven D. 2010. Population genomic sequencing of Coccidioides fungi reveals recent hybridization and transposon control. *Genome research*, **20**(7), 938–946.

Nguyen, Marcus, Ekstrom, Alex, Li, Xueqiong, & Yin, Yanbin. 2015b. HGT-Finder: A New Tool for Horizontal Gene Transfer Finding and Application to Aspergillus genomes. *Toxins*, **7**, 4035–4053.

Ni, Min, Feretzaki, Marianna, Sun, Sheng, Wang, Xuying, & Heitman, Joseph. 2011. Sex in Fungi. *Annual Review of Genetics*, **45**(1), 405–430.

Niehaus, Eva-Maria, Kim, Hee-Kyoung, Münsterkötter, Martin, Janevska, Slavica, Arndt, Birgit, Kalinina, Svetlana A., Houterman, Petra M., Ahn, Il-Pyung, Alberti, Ilaria, Tonti, Stefano, Kim, Da-Woon, Sieber, Christian M. K., Humpf, Hans-Ulrich, Yun, Sung-Hwan, Güldener, Ulrich, & Tudzynski, Bettina. 2017. Comparative genomics of geographically distant Fusarium fujikuroi isolates revealed two distinct pathotypes correlating with secondary metabolite profiles. *PLOS Pathogens*, **13**(10), e1006670.

Nielsen, Jens Christian, Grijseels, Sietske, Prigent, Sylvain, Ji, Boyang, Dainat, Jacques, Nielsen, Kristian Fog, Frisvad, Jens Christian, Workman, Mhairi, & Nielsen, Jens. 2017. Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in Penicillium species. *Nature Microbiology*, **2**(6), 17044.

Nijkamp, Jurgen F, van den Broek, Marcel, Datema, Erwin, de Kok, Stefan, Bosman, Lizanne, Luttik, Marijke A, Daran-Lapujade, Pascale, Vongsangnak, Wanwipa, Nielsen, Jens, Heijne, Wilbert H M, Klaassen, Paul, Paddon, Chris J, Platt, Darren, Kötter, Peter, van Ham, Roeland C, Reinders, Marcel J T, Pronk, Jack T, de Ridder, Dick, & Daran, Jean-Marc. 2012. De novo sequencing, assembly and analysis of the genome of the laboratory strain Saccharomyces cerevisiae CEN.PK113-7D, a model for modern industrial biotechnology. *Microbial cell factories*, **11**(mar), 36.

Nobre, Tania, Koopmanschap, Bertha, Baars, Johan J P, Sonnenberg, Anton S M, & Aanen, Duur K. 2014. The scope for nuclear selection within Termitomyces fungi associated with fungus-growing termites is limited. *BMC evolutionary biology*, **14**(1), 121.

Novo, Maite, Bigey, Frédéric, Beyne, Emmanuelle, Galeote, Virginie, Gavory, Frédérick, Mallet, Sandrine, Cambon, Brigitte, Legras, Jean-Luc, Wincker, Patrick, Casaregola, Serge, & Dequin, Sylvie. 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. *Proceedings of the National Academy of Sciences*, **106**(38), 16333–8.

Okagaki, Laura H, & Nielsen, Kirsten. 2012. Titan cells confer protection from phagocytosis in Cryptococcus neoformans infections. *Eukaryotic cell*, **11**(6), 820–6.

Oliver, Richard P, & Solomon, Peter S. 2008. Recent fungal diseases of crop plants: is lateral gene transfer a common theme? *Molecular plant-microbe interactions : MPMI*, **21**(3), 287–93.

Oromendia, Ana B, Dodgson, Stacie E, & Amon, Angelika. 2012. Aneuploidy causes proteotoxic stress in yeast. *Genes & development*, **26**(24), 2696–708.

Ostrosky-zeichner, Luis. 2012. Invasive Mycoses: Diagnostic Challenges. *AJM*, **125**(1), S14–S24.

Otto, Sarah P. 2007. The Evolutionary Consequences of Polyploidy. *Cell*, **131**(3), 452–462.

Park, Robert F, & Wellings, Colin R. 2012. Somatic hybridization in the Uredinales. *Annual review of phytopathology*, **50**(jan), 219–39.

Peberdy, J F. 1979. Fungal Protoplasts: Isolation, Reversion, and Fusion. *Annual Review of Microbiology*, **33**(1), 21–39.

Peberdy, John F. 1989. Presidential address: Fungi without coats — protoplasts as tools for mycological research. *Mycological Research*, **93**(1), ii–20.

Pelin, Adrian, Selman, Mohammed, Aris-Brosou, Stéphane, Farinelli, Laurent, & Corradi, Nicolas. 2015. Genome analyses suggest the presence of polyploidy and recent human-driven expansions in eight global populations of the honeybee pathogen Nosema ceranae. *Environmental Microbiology*, **17**(11).

Peter, Jackson, De Chiara, Matteo, Friedrich, Anne, Yue, Jia-Xing, Pflieger, David, Bergström, Anders, Sigwalt, Anastasie, Barre, Benjamin, Freel, Kelle, Llored, Agnès, Cruaud, Corinne, Labadie, Karine, Aury, Jean-Marc, Istace, Benjamin, Lebrigand, Kevin, Barbry, Pascal, Engelen, Stefan, Lemainque, Arnaud, Wincker, Patrick, Liti, Gianni, & Schacherer, Joseph. 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates Species-wide genetic and phenotypic diversity. *Nature*, **556**.

Plissonneau, Clémence, Stürchler, Alessandra, Croll, Daniel, & Taylor, John W. 2016. The Evolution of Orphan Regions in Genomes of a Fungal Pathogen of Wheat. *mBio*, **7**(5).

Plissonneau, Clémence, Hartmann, Fanny E., & Croll, Daniel. 2018. Pangenome analyses of the wheat pathogen Zymoseptoria tritici reveal the structural basis of a highly plastic eukaryotic genome. *BMC Biology*, **16**(1), 5.

Pontecorvo, G. 1956. The Parasexual Cycle in Fungi. *Annual Review of Microbiology*, **10**(1), 393–400.

Ponting, Chris P. 2017. Biological function in the twilight zone of sequence conservation. *BMC Biology*, **15**(1), 71.

Poppe, Stephan, Dorsheimer, Lena, Happel, Petra, & Stukenbrock, Eva Holtgrewe. 2015. Rapidly Evolving Genes Are Key Players in Host Specialization and Virulence of the Fungal Wheat Pathogen Zymoseptoria tritici (Mycosphaerella graminicola). *PLoS pathogens*, **11**(7).

Pryszcz, Leszek P, & Gabaldón, Toni. 2016. Redundans : an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Research*, **8**(44), 1–16.

Pryszcz, Leszek P, Nemeth, Tibor, Gacser, Atthila, & Gabaldon, Toni. 2014. Genome Comparison of Candida orthopsilosis Clinical Strains Reveals the Existence of Hybrids between Two Distinct Subspecies ´. *Genome biology and evolution*, **6**(5), 1069–1078.

Pryszcz, Leszek P., Németh, Tibor, Saus, Ester, Ksiezopolska, Ewa, Hegedűsová, Eva, Nosek, Jozef, Wolfe, Kenneth H., Gacser, Attila, & Gabaldón, Toni. 2015. The Genomic Aftermath of Hybridization in the Opportunistic Pathogen Candida metapsilosis. *PLOS Genetics*, **11**(10), e1005626.

Qiu, Huan, Cai, Guohong, Luo, Jing, Bhattacharya, Debashish, & Zhang, Ning. 2016. Extensive horizontal gene transfers between plant pathogenic fungi. *BMC biology*, **14**(1), 41.

Raudaskoski, Marjatta, & Kothe, Erika. 2010. Basidiomycete mating type genes and pheromone signaling. *Eukaryotic cell*, **9**(6), 847–59.

Richards, Thomas A, Dacks, Joel B, Jenkinson, Joanna M, Thornton, Christopher R, & Talbot, Nicholas J. 2006. Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. *Current biology*, **16**(18), 1857–64.

Richards, Thomas A., Leonard, Guy, Soanes, Darren M., & Talbot, Nicholas J. 2011a. Gene transfer into the fungi. *Fungal Biology Reviews*, **25**(2), 98–110.

Roberts, Samantha E, & Gladfelter, Amy S. 2015. Nuclear autonomy in multinucleate fungi. *Current Opinion in Microbiology*, **28**, 60–65.

Ropars, Jeanne, & Corradi, Nicolas. 2015. Homokaryotic vs heterokaryotic mycelium in arbuscular mycorrhizal fungi: different techniques, different results? *New Phytologist*, **208**(3), 638–641.

Ropars, Jeanne, Rodríguez de la Vega, Ricardo C., López-Villavicencio, Manuela, Gouzy, Jérôme, Sallet, Erika, Dumas, Émilie, Lacoste, Sandrine, Debuchy, Robert, Dupont, Joëlle, Branca, Antoine, & Giraud, Tatiana. 2015. Adaptive Horizontal Gene Transfers between Multiple Cheese-Associated Fungi. *Current Biology*, **25**, 1–8.

Roper, Marcus, Ellison, Chris, Taylor, John W, & Glass, N Louise. 2012. Nuclear and Genome Dynamics in Multinucleate Ascomycete Fungi. *Current Biology*, **21**(18), R786–93.

Roper, Marcus, Simonin, Anna, Hickey, Patrick C, Leeder, Abby, & Glass, N Louise. 2013. Nuclear dynamics in a fungal chimera. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(32), 12875–80.

Roper, Marcus, Lee, Changhwan, Hickey, Patrick C, & Gladfelter, Amy S. 2015. Life as a moving fluid: fate of cytoplasmic macromolecules in dynamic fungal syncytia. *Current Opinion in Microbiology*, **26**, 116–122.

Rosenblum, Erica Bree, James, Timothy Y, Zamudio, Kelly R, Poorten, Thomas J, Ilut, Dan, Rodriguez, David, Eastman, Jonathan M, Richards-Hrdlicka, Katy, Joneson, Suzanne, Jenkinson, Thomas S, Longcore, Joyce E, Parra Olea, Gabriela, Toledo, Luís Felipe, Arellano, Maria Luz, Medina, Edgar M, Restrepo, Silvia, Flechas, Sandra Victoria, Berger, Lee, Briggs, Cheryl J, & Stajich, Jason E. 2013a. Complex history of the amphibian-killing chytrid fungus revealed with genome resequencing data. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(23), 9385–90.

Rosenzweig, R Frank, Sharp, R R, Treves, David S, & Adams, Julian. 1994. Microbial Evolution in a Simple Unstructured Environment: Genetic Differentiation in Escherichia coli. *Genetics*, **137**, 903–917.

Rosewich, U. Liane, & Kistler, H. Corby. 2000. Role of Horizontal Gene Transfer in the Evolution of Fungi. *Annual Review Phytopathology*, **38**(325), 63.

Ryan, F J, & Lederberg, J. 1946. Reverse-Mutation and Adaptation in Leucineless Neurospora. *Proceedings of the National Academy of Sciences*, **32**(6), 163–73.

Ryan, Francis J., Beadle, G. W., & Tatum, E. L. 1943. The Tube Method of Measuring the Growth Rate of Neurospora. *American Journal of Botany*, **30**(10), 784.

Saari, S, & Faeth, S H. 2012. Hybridization of Neotyphodium endophytes enhances competitive ability of the host grass. *New phytologist*, **195**, 231–236.

Safonova, Yana, Bankevich, Anton, & Pevzner, Pavel A. 2015. dipSPAdes: Assembler for Highly Polymorphic Diploid Genomes. *Journal of computational biology : a journal of computational molecular cell biology*, **22**(6), 528–45.

Salzberg, Steven L. 2017. Horizontal gene transfer is not a hallmark of the human genome. *Genome Biology*, **18**(1), 85.

Sanz, Catalina, Alvarez, María I, Orejas, Margarita, Velayos, Antonio, Eslava, Arturo P, & Benito, Ernesto P. 2002. Interallelic complementation provides genetic evidence for the multimeric organization of the Phycomyces blakesleeanus phytoene dehydrogenase. *European journal of biochemistry*, **269**(3), 902–8.

Saupe, S J. 2000. Molecular genetics of heterokaryon incompatibility in filamentous ascomycetes. *Microbiology and molecular biology reviews*, **64**(3), 489–502.

Sawasaki, Y, Inomata, K, & Yoshida, K. 1996. Trans-kingdom conjugation between Agrobacterium tumefaciens and Saccharomyces cerevisiae, a bacterium and a yeast. *Plant & cell physiology*, **37**(1), 103–6.

Schmitt, Imke, & Lumbsch, H Thorsten. 2009. Ancient horizontal gene transfer from bacteria enhances biosynthetic capabilities of fungi. *PloS one*, **4**(2), e4437.

Schoenfelder, Kevin P, & Fox, Donald T. 2015. The expanding implications of polyploidy. *The Journal of cell biology*, **209**(4), 485–91.

Schwager, Evelyn E, Sharma, Prashant P, Clarke, Thomas, Leite, Daniel J, Wierschin, Torsten, Pechmann, Matthias, Akiyama-Oda, Yasuko, Esposito, Lauren, Bechsgaard, Jesper, Bilde, Trine, Buffry, Alexandra D, Chao, Hsu, Dinh, Huyen, Doddapaneni, HarshaVardhan, Dugan, Shannon, Eib-

ner, Cornelius, Extavour, Cassandra G, Funch, Peter, Garb, Jessica, Gonzalez, Luis B, Gonzalez, Vanessa L, Griffiths-Jones, Sam, Han, Yi, Hayashi, Cheryl, Hilbrant, Maarten, Hughes, Daniel S T, Janssen, Ralf, Lee, Sandra L, Maeso, Ignacio, Murali, Shwetha C, Muzny, Donna M, Nunes da Fonseca, Rodrigo, Paese, Christian L B, Qu, Jiaxin, Ronshaugen, Matthew, Schomburg, Christoph, Schönauer, Anna, Stollewerk, Angelika, Torres-Oliva, Montserrat, Turetzek, Natascha, Vanthournout, Bram, Werren, John H, Wolff, Carsten, Worley, Kim C, Bucher, Gregor, Gibbs, Richard A, Coddington, Jonathan, Oda, Hiroki, Stanke, Mario, Ayoub, Nadia A, Prpic, Nikola-Michael, Flot, Jean-François, Posnien, Nico, Richards, Stephen, & McGregor, Alistair P. 2017. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC biology*, **15**(1), 62.

Scott, Amber L, Richmond, Phillip A, Dowell, Robin D, & Selmecki, Anna M. 2017. The Influence of Polyploidy on the Evolution of Yeast Grown in a Sub-Optimal Carbon Source. *Molecular biology and evolution*, **34**(10), 2690–2703.

Shepherd, V A, Orlovich, D A, & Ashford, A E. 1993. Cell-to-cell transport via motile tubules in growing hyphae of a fungus. *Journal of cell science*, **105 (Pt 4**(aug), 1173–8.

Shoji, Jun-Ya, Charlton, Nikki D, Yi, Mihwa, Young, Carolyn A, & Craven, Kelly D. 2015. Vegetative hyphal fusion and subsequent nuclear behavior in Epichloë grass endophytes. *PloS one*, **10**(4).

Sinha, Sunita, Flibotte, Stephane, Niera, Mauricio, Formby, Sean, Plemenitaš, Ana, Gunde Cimerman, Nina, Lenassi, Metka, Gostinčar, Cene, Stajich, Jason E., & Nislow, Corey. 2017. Insight into the Recent Genome Duplication of the Halophilic Yeast Hortaea werneckii: Combining an Improved Genome with Gene Expression and Chromatin Structure. *G3: Genes, Genomes, Genetics*.

Sionov, Edward, Lee, Hyeseung, Chang, Yun C, & Kwon-Chung, Kyung J. 2010. Cryptococcus neoformans overcomes stress of azole drugs by formation of disomy in specific multiple chromosomes. *PLoS pathogens*, **6**(4), e1000848.

Sipos, György, Anderson, James B, & Nagy, László G. 2018. Armillaria. *Current biology*, **28**(7), R297–R298.

Slot, Jason C, & Hibbett, David S. 2007. Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study. *PloS one*, **2**(10), e1097.

Soanes, Darren, & Richards, Thomas A. 2014. Horizontal Gene Transfer in Eukaryotic Plant Pathogens. *Annual review of phytopathology*, **52**(aug), 583–614.

Soltis, Pamela S, Marchant, D Blaine, Van de Peer, Yves, & Soltis, Douglas E. 2015. Polyploidy and genome evolution in plants. *Current Opinion in Genetics & Development*, **35**(dec), 119–125.

Soucy, Shannon M., Huang, Jinling, & Gogarten, Johann Peter. 2015. Horizontal gene transfer: building the web of life. *Nature Reviews Genetics*, **16**(8), 472–482.

Springer, Nathan M, Ying, Kai, Fu, Yan, Ji, Tieming, Yeh, Cheng-Ting, Jia, Yi, Wu, Wei, Richmond, Todd, Kitzman, Jacob, Rosenbaum, Heidi, Iniguez, A Leonardo, Barbazuk, W Brad, Jeddeloh, Jeffrey A, Nettleton, Daniel, & Schnable, Patrick S. 2009. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS genetics*, **5**(11), e1000734.

Sriswasdi, Sira, Takashima, Masako, Manabe, Ri-Ichiroh, Ohkuma, Moriya, Sugita, Takashi, & Iwasaki, Wataru. 2016. Global deceleration of gene evolution following recent genome hybridizations in fungi. *Genome research*, **26**(8), 1081–90.

Stanhope, Michael J., Lupas, Andrei, Italia, Michael J., Koretke, Kristin K., Volker, Craig, & Brown, James R. 2001. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature*, **411**(6840), 940–944.

Strom, Noah B, & Bushley, Kathryn E. 2016. Two genomes are better than one: history, genetics, and biotechnological applications of fungal heterokaryons. *Fungal biology and biotechnology*, **3**, 4.

Strope, Pooja K, Skelly, Daniel A, Kozmin, Stanislav G, Mahadevan, Gayathri, Stone, Eric A, Magwene, Paul M, Dietrich, Fred S, & McCusker, John H. 2015a. The 100-genomes strains, an S. cerevisiae resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome research*, **25**(5), 762–74.

Stukenbrock, Eva H. 2016. The Role of Hybridization in the Evolution and Emergence of New Fungal Plant Pathogens. *Phytopathology*, **106**(2), 104–112.

Stukenbrock, Eva Holtgrewe, Bataillon, Thomas, Dutheil, Julien Y, Hansen, Troels T, Li, Ruiqiang, Zala, Marcello, Mcdonald, Bruce A, Wang, Jun, & Schierup, Mikkel H. 2011. The making of a new pathogen: Insights from comparative population genomics of the domesticated wheat pathogen Mycosphaerella graminicola and its wild sister species. *Genome research*, **21**, 2157–2166.

Stukenbrock, Eva Holtgrewe, Christiansen, Freddy Bugge, Hansen, Troels Toftebjerg, Dutheil, Julien Yann, & Schierup, Mikkel Heide. 2012. Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proceedings of the National Academy of Sciences*, **109**(27), 10954–10959.

Sun, Sheng, Billmyre, R. Blake, Mieczkowski, Piotr a., & Heitman, Joseph. 2014. Unisexual Reproduction Drives Meiotic Recombination and Phenotypic and Karyotypic Plasticity in Cryptococcus neoformans. *PLoS Genetics*, **10**(12), e1004849.

Suzuki, Katsunori, Moriguchi, Kazuki, & Yamamoto, Shinji. 2015. Horizontal DNA transfer from bacteria to eukaryotes and a lesson from experimental transfers. *Research in Microbiologoy*, **166**(10), 753–763.

Szöllősi, Gergely J., Davín, Adrián Arellano, Tannier, Eric, Daubin, Vincent, & Boussau, Bastien. 2015. Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **370**(1678).

Tetz, Victor V. 2005. The pangenome concept: a unifying view of genetic information. *Medical science monitor : international medical journal of experimental and clinical research*, **11**(7), HY24–9.

Todd, Robert T, Forche, Anja, & Selmecki, Anna. 2017. Ploidy Variation in Fungi: Polyploidy, Aneuploidy,

and Genome Evolution. *Microbiology spectrum*, **5**(4).

Torres, Eduardo M, Williams, Bret R, & Amon, Angelika. 2008. Aneuploidy: cells losing their balance. *Genetics*, **179**(2), 737–46.

Valero-Jiménez, Claudio A., Faino, Luigi, Spring in't Veld, Daphne, Smit, Sandra, Zwaan, Bas J., & van Kan, Jan A. L. 2016. Comparative genomics of Beauveria bassiana: uncovering signatures of virulence against mosquitoes. *BMC Genomics*, **17**(1), 986.

Van de Peer, Yves. 2004. Tetraodon genome confirms Takifugu findings: most fish are ancient polyploids. *Genome biology*, **5**(12), 250.

Van de Peer, Yves, Taylor, John S, & Meyer, Axel. 2003. Are all fishes ancient polyploids? *Journal of structural and functional genomics*, **3**(1-4), 65–73.

Van Der Does, H. Charlotte, & Rep, Martijn. 2012. Horizontal transfer of supernumerary chromosomes in fungi. *Methods in Molecular Biology*, **835**, 427–437.

Van der Nest, M.A., Olson, Å., Lind, M., Vélëz, H., Dalman, K., Durling, M. Brandström, Karlsson, M., & Stenlid, J. 2014. Distribution and evolution of het gene homologs in the basidiomycota. *Fungal Genetics and Biology*, **64**(mar), 45–57.

Walther, Andrea, Hesselbart, Ana, & Wendland, Jürgen. 2014. Genome Sequence of Saccharomyces carlsbergensis, the World's First Pure Culture Lager Yeast. *G3: Genes, Genomes, Genetics*, **4**(5), 1–11.

Weber, T, & Marahiel, M a. 2001. Exploring the domain structure of modular nonribosomal peptide synthetases. *Structure*, **9**(1), R3–9.

Whelan, William L., Markie, David M., Simpkin, K. Graham, & Poulter, Russell M. 1985. Instability of Candida albicans. *Journal of bacteriology*, **161**(3), 1131–1136.

Wisecaver, Jennifer H., & Rokas, Antonis. 2015. Fungal metabolic gene clusters: caravans traveling across genomes and environments. *Frontiers in Microbiology*, **6**(mar), 161.

Wisecaver, Jennifer H., Slot, Jason C., & Rokas, Antonis. 2014. The Evolution of Fungal Metabolic Pathways. *PLoS Genet*, **10**(12), e1004816.

Wolfe, Kenneth H., & Shields, Denis C. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, **387**(6634), 708–713.

Wortel, Meike T., Bosdriesz, Evert, Teusink, Bas, & Bruggeman, Frank J. 2016. Evolutionary pressures on microbial metabolic strategies in the chemostat. *Scientific Reports*, **6**(1), 29503.

Wyss, Tania, Masclaux, Frédéric G, Rosikiewicz, Pawel, Pagni, Marco, & Sanders, Ian R. 2016. Population genomics reveals that within-fungus polymorphism is common and maintained in populations of the mycorrhizal fungus Rhizophagus irregularis. *The ISME journal*, **10**(10), 2514–26.

Yin, Zhiyuan, Zhu, Baitao, Feng, Hao, & Huang, Lili. 2016. Horizontal gene transfer drives adaptive colonization of apple trees by the fungal pathogen Valsa mali. *Scientific reports*, **6**(sep), 33129.

Zeyl, Clifford. 2009. The role of sex in fungal evolution. *Current Opinion in Microbiology*, **12**(6), 592–598.

Zhan, J, Pettway, R E, & McDonald, B A. 2003. The global genetic structure of the wheat pathogen Mycosphaerella graminicola is characterized by high nuclear diversity, low mitochondrial diversity, regular recombination, and gene flow. *Fungal genetics and biology*, **38**(3), 286–97.

Zhu, Yuan O., Sherlock, Gavin, & Petrov, Dmitri A. 2016. Whole Genome Analysis of 132 Clinical Saccharomyces cerevisiae Strains Reveals Extensive Ploidy Variation. *G3: Genes, Genomes, Genetics*, **6**(8).

Chapter 4

Widespread inter- and intra-domain horizontal gene transfer of D-amino acid metabolism enzymes in eukaryotes

Naranjo-Ortiz, M. A. et al. Widespread Inter- and Intra-Domain Horizontal Gene Transfer of d-Amino Acid Metabolism Enzymes in Eukaryotes. Front. Microbiol. **7**, 2001 (2016). URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5169069/ doi: 10.3389/fmicb.2016.02001

Naranjo-Ortíz MA, Brock M, Brunke S, Hube B, Marcet-Houben M, Gabaldón T. Widespread Inter- and Intra-Domain Horizontal Gene Transfer of d-Amino Acid Metabolism Enzymes in Eukaryotes. Front Microbiol. 2016 Dec 20;7. DOI: 10.3389/fmicb.2016.02001

Chapter 5

Bioinformatic analyses of unorthodox fungal genomes

Miguel Á. Naranjo-Ortiz, Toni Gabaldón

Abstract

Recent developments in sequencing technologies and bioinformatics tools have made genome sequencing and assembly accessible to many groups. However, the process of genome assembly can be challenging due to some intrinsic properties of the genome such as high heterozygosity, or the presence of polyploidies, aneuploidies, or heterokaryosis. Due to lack of expertise or appropriate tools, many such cases are likely to go unnoticed, potentially resulting in low quality assemblies being nevertheless deposited in public databases. Here we present Karyon, a python based toolkit that aims to help researches to understand the overall architecture of a *de novo* genome assembly. Karyon generates a set of plots that can be used to understand the genomic structure of a problematic organism, which in turn might provide an interesting glimpse into its biology and guide researchers towards the use of more specialized approaches. We include an automatized standard *de novo* genome assembly based on Redundans and a variant calling pipeline for its use as a complement to Karyon. Based on this, we hypothesized that genome databases must contain a non-negligible fraction of low quality assemblies that result from such type of intrinsic genomic factors. To test this, and quantify the different types of naturally occurring non-canonical genomic organizations, we used an automated genome diagnosis pipeline (Karyon) to assess 35 genome assemblies from 19 different Mucorales species deposited in Genebank NCBI database between 2005 and 2015. Our results show that 6 (17%) of the assemblies presented signs of unusual genomic configurations, suggesting that these events are common in Fungi and that they should be taken into consideration in genome assembly projects.

5.1 Karyon: Toolkit for unorthodox *de novo* genome assemblies

5.1.1 Introduction

Genomics have revolutionized biological research. A genome sequence offers a window to an organism's biology and constitutes an invaluable resource for research. Recent developments in sequencing technologies and software tools have made genome sequencing accessible to most research groups worldwide, which results in an exponential increase of genome sequences from a growing diversity of organisms, including unculturable ones. Despite recent developments and the availability of a plethora of tools, genome assembly is often a complex task that requires expertise. A successful genome assembly project also depends on the quality, design, and depth of the sequencing libraries which are often compromised in projects with restricted budget. Moreover intrinsic genome properties such as size, amount of repetitive sequences, or levels of heterozygosity, can make the process of genome assembly a challenging task, as the properties of the genome sequence deviate from default expectations of the assembly algorithms. For instance, the presence of high levels of heterozygosity have traditionally limited genomic projects, often requiring out-of-the-box approaches to overcome it (Neale *et al.* , 2014; Session *et al.* , 2016a). Yet, high heterozygosity is common in many organisms, and reaches extreme levels in hybrids (Otto, 2007; Todd *et al.* , 2017). The recognition of high levels of heterozygosity distributed unevenly along the genome in some fungal genomes, which results from a process of hybridization followed by loss of heterozygosity, led to the development of Redundans (Pryszcz & Gabaldón, 2016) an assembly pipeline that includes a reduction step that eliminates uncollapsed contigs and allows the reconstruction of a contiguous, yet chimeric, assembly. Analogously, abnormal ploidy characteristics such as polyploidy and aneuploidy are common in many groups of eukaryotes (Schoenfelder & Fox, 2015; Torres *et al.* , 2008). These chromosomal aberrations have profound consequences for the biology of the organisms, ranging from fruit characteristics in crops (Benny-Byfield & Wendel, 2014; Salman-Minkov *et al.* , 2016) to adaptation to drugs in pathogens (Anderson *et al.* , 2017). Furthermore some organisms present poly-nucleated cells, with different nuclei perhaps presenting a different genetic makeup (Strom & Bushley, 2016). Finally, most genome sequencing projects undertake the sequencing of a bulk of cells, which can be highly heterogeneous and present contamination, particularly in the case of symbiotic, commensal or parasitic associations that can be inadvertedly present in the sample (Wijayawardena *et al.* , 2013). Knowing such properties before the start of the assembly process is useful, as it can help choosing the appropriate tools and parameters. However, nowadays, for most newly sequenced organisms, this information is not available. For these cases the analysis of sequencing libraries and preliminary assemblies can provide important hints into the genome complexity of an organisms.

Based on this principle we have developed Karyon, a python-based toolkit that aims to aid at the diagnosis of low quality *de novo* genome assemblies. Here we provide an overview of the functionalities of the program, as well as some recommendations for interpreting the results.

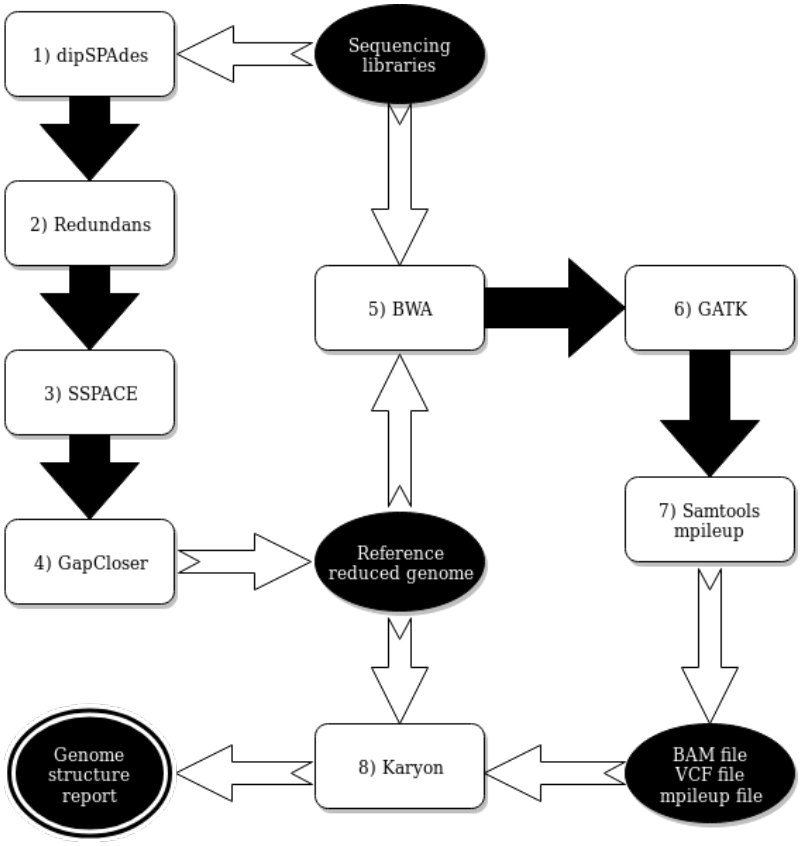

Figure 1: Karyon pipeline flowchart

Black circles and white arrows represent input files; white squares and black arrows represent consecutive computational steps.

5.1.2 Results

While Karyon can be used as a standalone program, it requires a previous variant calling protocol and we recommend its use over an assembly that has been reduced using Redundans. Karyon includes a wrapper that launches an assembly and variant calling pipeline on top of the standalone report program. This pipeline includes a fully functional all-around genome assembly protocol, and we encourage novice users to use the full pipeline. This assembly pipeline uses dipSPAdes as assembly protocol, a program that has been designed with bacterial and fungal genomes in mind and thus it might not be adequate for very large genomes. The pipeline also assumes the use of at least one Illumina paired-end sequencing library, and due to this and the selected genome assembler we recommend the use of other genome assemblers if other sequencing technologies are to be used.

Our pipeline evaluates library quality using FastQC and we filtered them using Trimmomatic (Bolger *et al.* , 2014). After filtering, we selected the library with deepest sequencing coverage and use it for the genome assembly using DipSPAdes (Safonova *et al.* , 2015), with default settings plus the tags "--expect_rearrangements" and "--expect-gaps" turned on. The resulting assembly is then used as input for Redundans (Pryszcz & Gabaldón, 2016) to generate and artificially reduced assembly, using default settings and all the paired sequencing libraries available. Redundans assembly was then used as a reference for mapping the deepest library using BWA (Li & Durbin, 2009). After this, it calls SNP using GATK (McKenna *et al.* , 2010). Finally, it generates the mpileup file using SAMtools (Li *et al.* , 2009a).

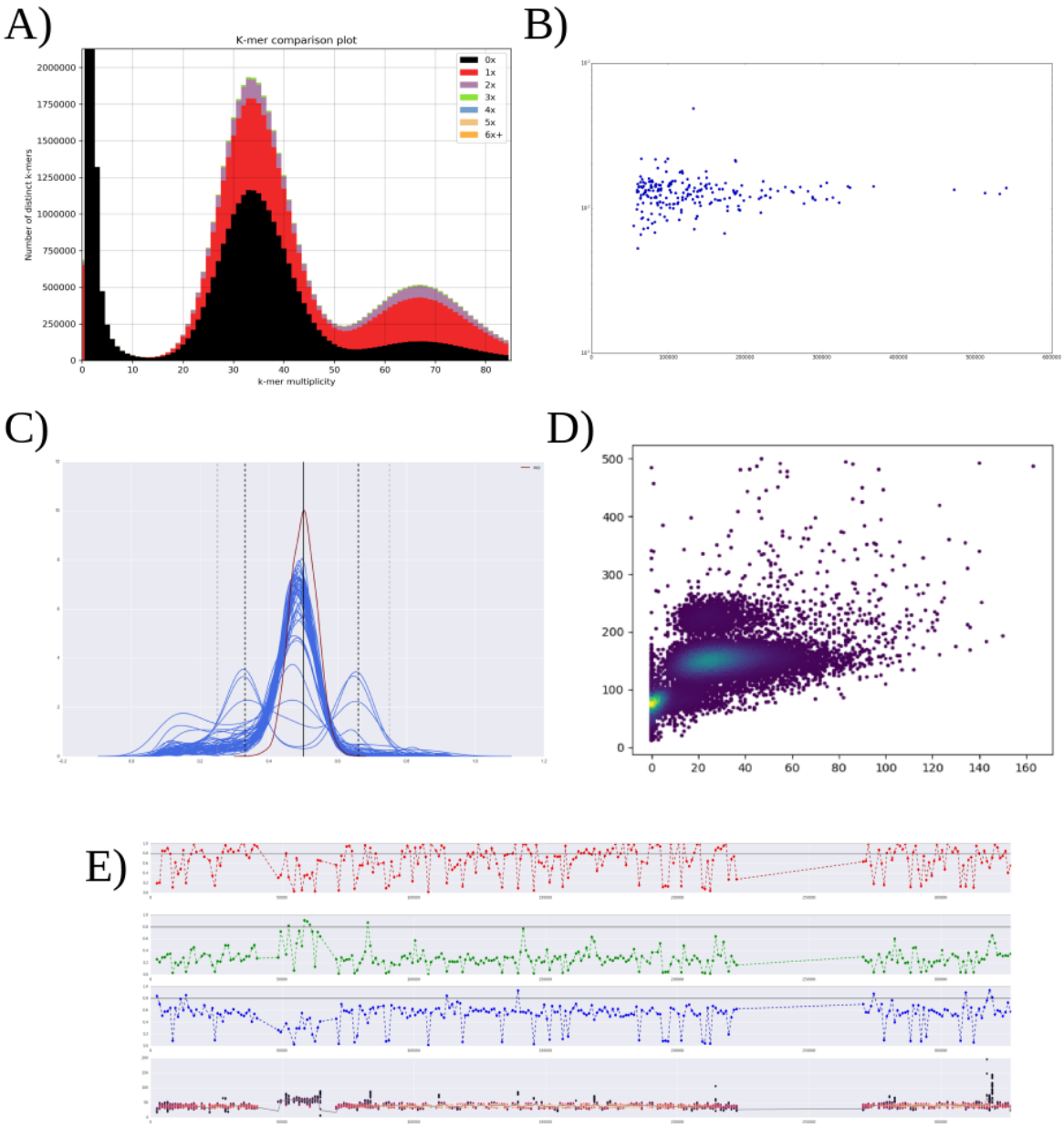

Figure 2. Karyon report example

A) *K*-mer distribution analysis performed with KAT (Mapleson *et al.* , 2016). Black area corresponds mostly with collapsed regions during the assembly using Redundans (Pryszcz & Gabaldón, 2016). **B)** Scatter plot of scaffold length (x axis) versus average coverage in said scaffold (y axis). **C)** Alternative allele frequency plot. Each blue line represents one scaffold, red line represents a simulated scaffold that follows a theoretical ideal diploid proportion. In this case, most scaffolds seem to behave similar to the theoretical diploid, but some tend to adjust to a theoretical triploid (thick dashed line). **D)** Variation versus coverage plot. Each dot represents a 1Kbp window of the genome whose coordinates correspond to the number of SNPs (x axis) and its average coverage (y axis). In the example we can observe a bimodal distribution of the dots, corresponding to a region with low coverage and low heterozygosity; and a region with high coverage and high heterozygosity. **D)** Local ploidy estimation using nQuire (Weiß *et al.* , 2017). The plot covers the length of a whole scaffold. The first three subplots represent the estimated probability of diploid, triploid and tetraploid model normalized by the corresponding free model per window of 1Kbp. The last subplot represents, per analyzed analyzed window, the coverage of each SNP site present in said window. Karyon generates automatically this plot for each scaffold.

Karyon will then construct the following plots:

Scaffold size distribution: Karyon plots sorted scaffolds versus their length in base pairs. Since The program uses a user defined window size for many analyses, this value appears as a dotted line in the plot, showcasing the fraction of scaffolds that would be affected by the selected size. Karyon produces two plots of this kind, one that uses linear representation of scaffolds length and a second that uses a logarithmic scale.

KAT *k*-mer analysis: Karyon uses KAT (Mapleson *et al.* , 2016) to provide a *k*-mer spectrum analysis as part of the whole report. This type of plots consist of a frequency histogram representing coverage in the horizontal axis and *k*-mer numbers in the vertical axis. In a haploid genome, for *k*-mers of enough size, most *k*-mers will appear either one or zero times. Those *k*-mers that appear once will have an average coverage roughly equal to the average depth of coverage. In the case of a non-homozygous diploid, most *k*-mers will be represented zero times, once or twice across the genome, and thus the *k*-mer plot should reflect two peaks.

Scaffold length versus coverage: Another plot that Karyon produces is a simple scatter plot showing the size of a contig versus the average sequencing depth of that contig. For most genomes, average scaffold coverage should be fairly uniform and dependent on the ploidy level. Aneuploidies should be observable as sub-populations of scaffolds with a fairly stable coverage that is nonetheless different from the overall coverage; or in case of well-resolved assemblies, as very long scaffolds with a drastic difference in coverage from the overall trend.

Alternative allele frequency plots: Karyon produces a frequency histogram of the number of reads mapped to an alternative SNP divided by coverage on the reference SNP in the same site. This is done for the genome as a whole and for a determined set of scaffolds independently, as two separated plots. In the second case, the program will sort scaffolds by length and represent the longest. For comparison, a simulated histogram of a perfect theoretical diploid is shown. This simulation is generated assuming the same number of sites with the same coverage of the sample, but assigning each read to reference or alternative randomly (fair coin), and should present a clear peak at 0.5. In the case of a diploid, the real data should adjust to the shape of the simulated histogram. In the case of a triploid, the histogram should present two asymmetrical peaks at 0.33 and 0.66. Peaks at very low values might be indicative of repetitive elements or of contaminating sequences, as these types of sequences should not follow the same apparent ploidy as the overall genome. Since the plot is also done for independent scaffolds, it can be used to identify aneuploidies.

Variation versus coverage: Karyon produces a scatter plot where each dot represents a region of the genome of a set window size. For each genomic window the plot represents the number of SNPs and the average sequencing coverage. A haploid genome will present most of its dots within a narrow cloud around zero SNP density and an average coverage roughly equal to the average genome coverage. A diploid genome will probably behave the same, but with a dot cloud shifted to the right in proportion to its heterozygosity. In this regard, certain highly heterozygous genomes present a secondary cloud with a normal average coverage but very low heterozygosity, which corresponds to loss of heterozygosity areas. Aneuploidies and haplome specific copy number variation would appear as dot clouds that move up or down the main trend.

Coverage plot: Karyon produces a box diagram showcasing the average coverage per window of user determined size for the 100 longest scaffolds. Since the scaffolds are ordered by size, it allows to identify if observed dual behavior correspond to poorly assembled scaffolds. Analogously, weird behaviors might be caused by many independent regions within a chromosome

Local ploidy estimation: Karyon uses nQuire (Weiß *et al.* , 2017) to estimate likelihood of diploid, triploid and tetraploid ploidy across windows of user defined size on each scaffold. The output is a plot per scaffold showcasing the distribution of SNPs across the sequence, as well as normalized likelihood of each ploidy. Normalized values are obtained by dividing the score of each ploidy level by the score of the free model. Since nQuire requires heterozygous data, low heterozygosity regions cannot obtain these scores and appear as blank areas in the plot. Nonetheless, this plot can complement the others and help to identify aneuploid scaffolds, duplicated regions and other other genomic abnormalities.

Availability

Karyon is written in Python 2.7 and freely available to download in:

https://github.com/Gabaldonlab/projects/tree/master/mnaranjo/karyon_pipeline

5.2 Non-canonical architectures in published Mucorales genomes

5.2.1 Introduction

Recent developments in high-throughput sequencing and computational pipelines for genome assembly an annotation have made the sequencing of a new organism, particularly microorganisms, a routine task for many laboratories. The sequencing of the full genome of an organism provides an invaluable resource to understand its biology at many different levels, from the molecular pathways that govern gene expression to the population structure of the species. Pioneering projects focused on model organisms and a few chosen species with great medical and economical relevance (Weissenbach, 2016). It soon became obvious that the difficulty of these projects was highly variable, and thus the type of data we were gathering was highly biased toward highly reduced and easy to sequence organisms. This greatly biases our comprehension of the natural world and, even worse, deliberately ignores a myriad of biological traits that are highly relevant for the physiology, ecology and evolution of many organisms. In recent years, the application of genomics approaches have expanded to virtually all types of organisms, including unculturable ones.

The success of a genome assembly project is limited by technical aspects as well as by intrinsic properties of the sequenced genome (Gabaldón & Alioto, 2016). Naturally, if the methodology used to sequence the sample is insufficient, the results would be poor. On the other hand, intrinsic factors are properties of the genome itself that interfere with genome assembly algorithms, thus resulting in low quality outputs. If the presence of challenging intrinsic factors is anticipated, specific approaches can be used in some cases. However, this generally increases the overall costs of a project and requires expertise that is not universally available. Often, such challenging factors are unanticipated and the possible causes for a low quality assembly are not investigated. As a result it is expected that many low quality assemblies exist that originated from undiscovered intrinsic features of the genome. A fraction of these assemblies might have been deposited, making it possible to use computational approaches on available genomes to diagnose underlying non-canonical genomic features.

The main intrinsic factors that compromise the success of a genome project are the genome size, the heterozygosity, the presence of low complexity regions, the changes in ploidy, and the presence of genomic contamination. Genome size greatly affects computational costs, as many of the algorithms used in these projects scale non-linearly (Simpson & Pop, 2015; Wajid & Serpedin, 2012; Wajid *et al.* , 2016). Heterozygosity is the presence of allelic differences within an individual, which can be single nucleotide polymorphisms, insertion-deletion differences or even gross genomic rearrangements. For regions in which heterozygous sites are abundant, it might be difficult to differentiate between a heterozygous region, or two independent albeit highly similar regions (Hirsch & Robin Buell, 2013; Pryszcz & Gabaldón, 2016). This in turn results in fragmented assemblies with an inflated total size, as many of these regions appear duplicated and poorly assembled. Many lineages contain genomes in which repetitive regions are abundant. Since sequencing reads are typically very short, if the repetitive region is longer that the read it can be virtually impossible to determine its length. The presence of several similar regions adds an additional layer of uncertainty, as it generally produces fragmented assemblies and greatly increases the computational costs (Hirsch & Robin Buell, 2013; Wajid *et al.* , 2016). Similarly, ploidy deviations play a very important role in genome assembly. The first possible deviation is polyploidy, which is the presence of a number of chromosome sets higher than two. Polyploidy is generally associated to genome

heterozygosity, as it increases the amount of possible states per site (Aguiar & Istrail, 2013; Bonizzoni *et al.* , 2016). Polyploid genomes risk suffering inappropriate recombinations during meiosis and mitosis, and thus tend to recover a diploid or diploid-like state (Otto, 2007; Schoenfelder & Fox, 2015). Highly divergent sub-genomes might stabilize by duplicating themselves, increasing the ploidy (allopolyploidy) and forming recombinant pairs with their own copies. In such cases, the parental sub-genomes act independently from each other, effectively behaving as a diploid with an increased set of chromosomes. Another possibility is to produce genomic rearrangements that stabilize chromosomes, occasionally causing regions of one of the sub genomes to be lost and be substituted by a copy of the other (gene conversion or loss of heterozygosity). This second option is more common in organisms that can reproduce asexually, since mitosis is much more tolerant toward chromosomic impairments than meiosis. Polyploidy can be inferred either experimentally or by analyzing sequencing libraries prior to genome assembly (Audano & Vannberg, 2014; Mamun *et al.* , 2016; Mapleson *et al.* , 2016; Marcais & Kingsford, 2012), although other related phenomena might complicate this task. The second deviation from the traditional eukaryotic karyotypic organization is aneuploidy, the presence of abnormal numbers of individual chromosomes (Gerstein & Berman, 2015; Torres *et al.* , 2008). Aneuploidy tends to cause the same problems as polyploidy, albeit with the effect being limited only to a portion of the genome, corresponding with the aneuploid chromosomes. Genomic contamination can normally be prevented by methodological means. However, in certain occasions the contaminating sequence is intrinsic for the organism, such as the case of endosymbiotic organisms. Finally, in syncytial organisms, such as filamentous fungi or slime moulds, there is the possibility of the coexistence within a cytoplasm of genetically different populations, a condition known as heterokaryosis (James *et al.* , 2008; Maheshwari, 2005; Strom & Bushley, 2016).

Since all of these scenarios affect genome assembly, it is easy to conclude that our genomic data is particularly biased against organisms with any of these conditions. On the other hand, the existence of some of these singularities in sequenced organisms might have been passed unnoticed to many researchers, as their detection is far from simple. It is so far unknown how common such non-canonical genome organizations are, and whether they can be recognized among the set of already sequenced genomes. Fungi are in a particularly privileged position to start addressing these questions. Fungal organisms usually have small and compact genomes and many of them can be cultured easily in axenic culture. Fungal genes tend to have a low intronic density, which eases the application of *de novo* annotation strategies. Because all of this, the amount of fungal genomes is now in the order of thousands, including several strains for many species, and the number keeps increasing. Many fungal species are sequenced due to their clinical, agricultural, ecological, or industrial importance, but comprehensive efforts to obtain a balanced coverage of the existing fungal diversity are ongoing such as the 1000 fungal genomes, and the 1000 yeast genomes initiative (Peter *et al.* , 2018; Strope *et al.* , 2015a; Wilkening *et al.* , 2013; Zhu *et al.* , 2016). Thus, fungi provide an excellent system to study the incidence of different genomic accidents in evolution (Berman *et al.* , 2016; Gerstein & Berman, 2015; Todd *et al.* , 2017). However, the quality of fungal genomes is often sub-optimal, and databases are riffed with highly fragmented assemblies. Many fungi present really thick cell walls, and synthesize protective pigments and other secondary metabolites that interfere with the processes of DNA extraction or even DNA library preparation (Kim *et al.* , 2010; Sreenivasaprasad, 2000). These mishaps can probably affect the cost efficiency of a sequencing project and have an incidence on the final results. Beyond biochemical limitations, genomic factors such as those discussed above may complicate genome assembly. Many small laboratories embark in the task of sequencing a fungal genome due to the relative low cost, but quite often they operate under budget limitations or employ personnel with limited bioinformatics training that might lack the expertise needed to identify and deal with unexpected situations. The existence of non-canonical genomic organizations as those described above are clearly a factor that increases the chances of obtaining unsatisfactory assemblies. Considering this, we hypothesize that genome databases must contain a fraction of low quality assemblies from fungal organisms that are caused by intrinsic genomic factors. If that is true, reanalysis of the raw data from low quality assemblies should lead us to describe novel genomic accidents. Here we test this idea on 34 deposited genomes from the fungal order Mucorales. Our results suggest that non-canonical genomic organizations are not rare, and that future studies on other fungal groups are likely to uncover many new cases.

5.2.2 Material and methods

Sequencing data: For all groups considered, we downloaded raw data from libraries deposited at Short Read Archive (SRA) (National Centre for Biotechnology Information, 2015) of those species presenting an highly fragmented assembly (>1000 scaffolds), which presented at least one paired-end Illumina library larger than 1Gb (after quality filtering), to ensure a minimal coverage.

Data analyses: For each species, we evaluated library quality using FastQC (Andrews, 2010) and we filtered them using Trimmomatic (Bolger *et al.* , 2014) After filtering, we selected the library with deepest sequencing coverage and use it for the genome assembly using DipSPAdes (Safonova *et al.* , 2015), with default settings plus the tags "--expect_rearrangements" and "--expect-gaps" turned on. The resulting assembly was then used as input for Redundans to generate an artificially reduced assembly, using default settings and all the paired sequencing libraries available. The redundans assembly was then used as a reference for mapping the deepest library using BWA (Li & Durbin, 2009). After this, we called SNP using GATK (McKenna *et al.* , 2010). This step is important to identify internal variability of the sample, such as differences in coverage or heterozygosity.

Contamination detection: For each of the suspicious genomes, we used the assembly to perform an Agustus (Brudno *et al.* , 2003) gene prediction. Then, we used Blast (Altschul, Stephen F and Gish, Warren and Miller, Webb & Lipman, 1990) to align the whole proteome against Uniref100 (Consortium, 2014). Since the genomes come from public databases, their own proteins should appear as hits and thus we selected the 10 best hits. We have used this hits to assign a taxonomic profile. Additionally, we have used the predicted Augustus CDS to map sequencing reads with GATK. With both the taxonomic profile and the variant calling file, we have run Blobtools (Laetsch & Blaxter, 2017) in order to identify the pressence of widespread contamination in the sequencing libraries.

Table 1: Analyzed genomes in the Mucorales

Species	NCBI genome size (Mbp)	NCBI number of scaffolds
Rhizopus microsporus ATCC62417	49.6	1386
Rhizopus microsporus CBS344.29	49.2	1554
Rhizopus microsporus B9738	75.13	5266
Rhizopus microsporus var. *rhizopodiformus* B7455	48.7	4658
Rhizopus delemar Type I NRRL 21789	42	3921
Rhizopus delemar Type II NRRL 21446	38.9	1156
Rhizopus delemar Type II NRRL 21447	38.7	1177
Rhizopus delemar Type II NRRL 21477	40.8	1808
Rhizopus oryzae 99-892	39.1	1168
Rhizopus oryzae HUMC02	40.3	2313
Rhizopus oryzae B7407	43.3	4683
Rhizopus oryzae type I NRRL 13440	43.4	5022
Rhizopus oryzae type I NRRL 18148	47.5	14653
Rhizopus oryzae type I NRRL 21396	42.8	4445
Rhizopus oryzae 99-133	41.5	4317
Rhizopus oryzae 97-1192	42.9	4566
Rhizopus stolonifer B9770	38	5567
Mucor circinelloides B8987	36.7	2210
Mucor indicus B7402	39.8	3117
Mucor racemosus B9645	65.5	6360
Mucor velutinosus B5328	35.9	2411
Lichtheimia corymbifera 008-049	36.6	1629
Lichtheimia corymbifera B2541	36.6	1176
Lichtheimia ramosa B5399	45.6	3968
Saksenaea oblongispora B3353	40.8	1702
Saksenaea vasiformis B4078	42.5	2417
Cokeromyces recurvatus B5483	29.3	2637
Syncephalastrum monosporum B8922	29.6	1284
Syncephalastrum racemosum B6101	29.6	1035
Cunninghamella elegans B9769	31.7	1380
Apophysomyces elegans B7760	38.5	1528
Apophysomyces trapeziformis B9324	35.8	1400
Thermomucor indicae-seudaticae HACC243	29.55	1958
Parasitella parasitica CBS 412.66 isolate NGI315 ade- mutant	44.9	15637

5.2.3 Results

We selected the order Mucorales because this group comprises several described examples of whole genome duplication, both at ancient and recent levels (Corrochano *et al.* , 2016; Ma *et al.* , 2009). Many sequenced members of the clade come from clinical samples, an environment that is known to promote the emergence of different genomic accidents (Schoenfelder & Fox, 2015; Todd *et al.* , 2017). Additionally, several represented species present two or more sequenced isolates, allowing to get a glimpse at their intra-specific diversity. We obtained 35 genome assemblies from 19 different Mucorales species deposited in Genebank between January 1st 2005 and December 31th 2015 (Table 1). We re-assembled these genomes with Redundans (Pryszcz & Gabaldón, 2016), and used the Karyon pipeline to diagnose potential non-canonical genome organizations. Most of the analyzed genomes (27, 79.4%) presented very low levels of heterozygosity and a relatively homogeneous coverage across the genome, suggesting that those strains are haploid or, if presenting higher ploidy, extremely homozygous. However, our pipeline uncovered cases that produced anomalous

results in the different karyon tests. Below we describe these cases and propose a plausible scenario to explain each of the obtained results.

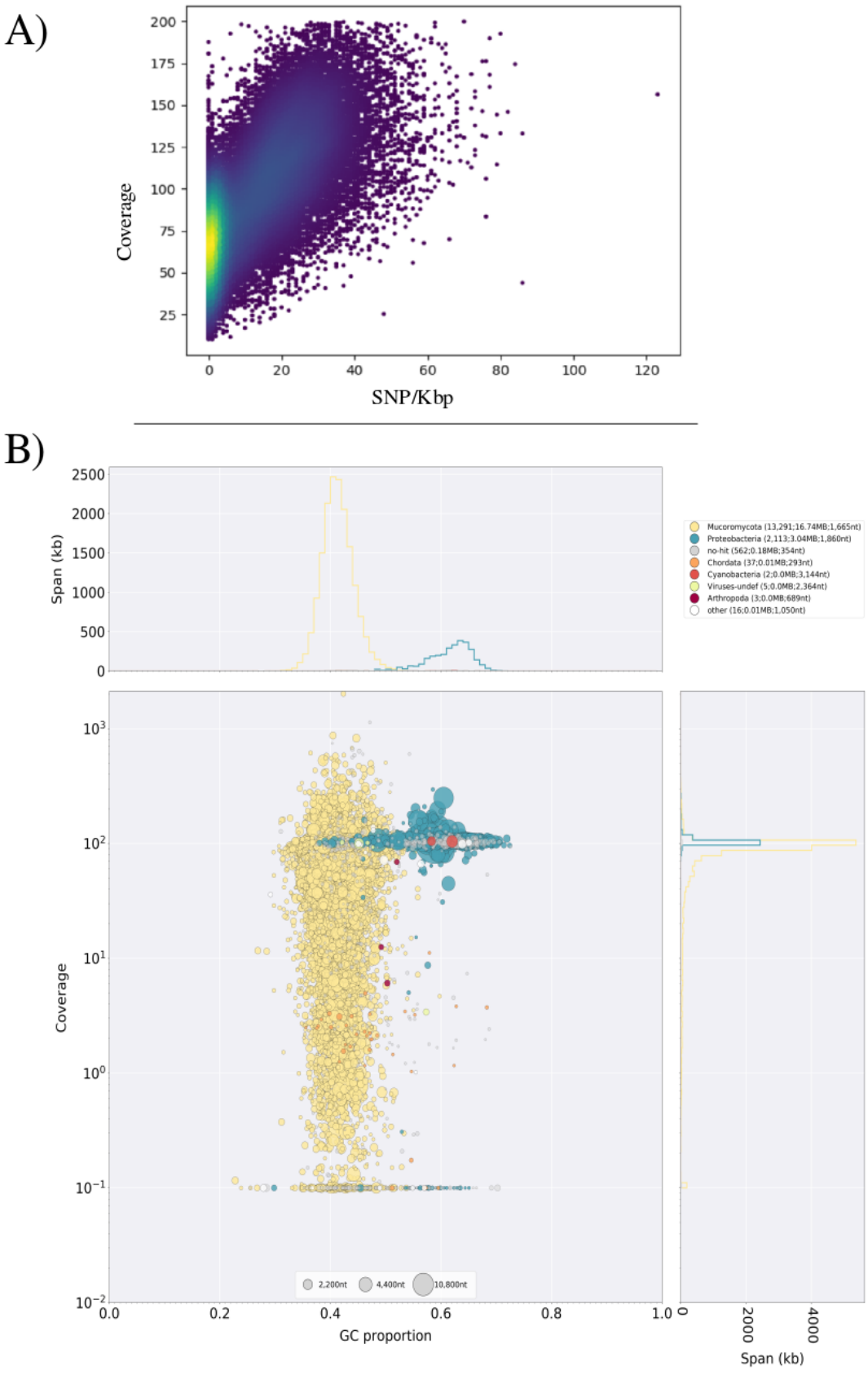

Figure 3. Analysis of *Rhizopus microsporus* ATCC62417
A) Variation versus coverage plot reveals the existence of a highly variable portion of the genome that presents very variable heterozygosity levels. **B)** Blobtools analyses suggest that the genome presents a considerable portion of contaminating sequences. Coverage of the sequences assigned to bacteria is very low when the analyses are performed with other libraries (Data not shown), which proves that the conflicting signal has its origin in a contaminated sequencing library. Results for *R. microsporus* and CBS344.5 and var. *rhizopodiformus* B7455 are very similar (data not shown).

At the moment of this study, eight *Rhizopus microsporus* strains were deposited in the NCBI database. Interestingly, three of them presented a genome size estimated around 25Mbp; four of them had a genome size close to 50Mbp; and one presented a genome size of 75Mbp. The three strains with a genome size of 25Mbp had sufficiently good assemblies with a scaffold number below 1000, and thus were not selected for further analyses. Additionally, the raw libraries for one of the strains (*Rhizopus microsporus* var. *chinensis* CCTCC M201021) were not publicly available and thus could not be part of the survey. For three of the analyzed strains (ATCC62417, CBS344.5 and var *rhizopodiformus* B7455) our *de novo* assembly pipeline recovered a genome size of approximately 40Mb, which is smaller than the assemblies deposited in NCBI. The heterozygosity distribution in these assemblies shows that most of the genome presented a relatively uniform behavior with low heterozygosity. In all three cases, though, a considerable proportion of the genome appears with a highly variable coverage and increased heterozygosity (Figure 3). These regions seem, in all three cases, to present a higher coverage than the main fraction of the genome. The remaining strain B9738 was a big surprise. This strain, that has a genome size of 75Mbp in NCBI and 71Mbp in our assembly; presents an extremely low level heterozygosity and a very homogeneous coverage. *K*-mer spectrum also shows just one very clear peak. All in all, all this suggests that B9738 is haploid, despite presenting a 3-fold increase in genome size compared to other strains of the same species (Figure 4). Gene prediction with Augustus, using *Rhizopus oryzae* as trained model, returned a total of 21,300 gene models, which is very high for a fungus. Contamination analyses does not suggest the presence of widespread contamination that could explain such over-inflated genome (Figure 4). We suggest that B9738 might be a misidentified strain and that additional analyses should be performed to clarify this situation.

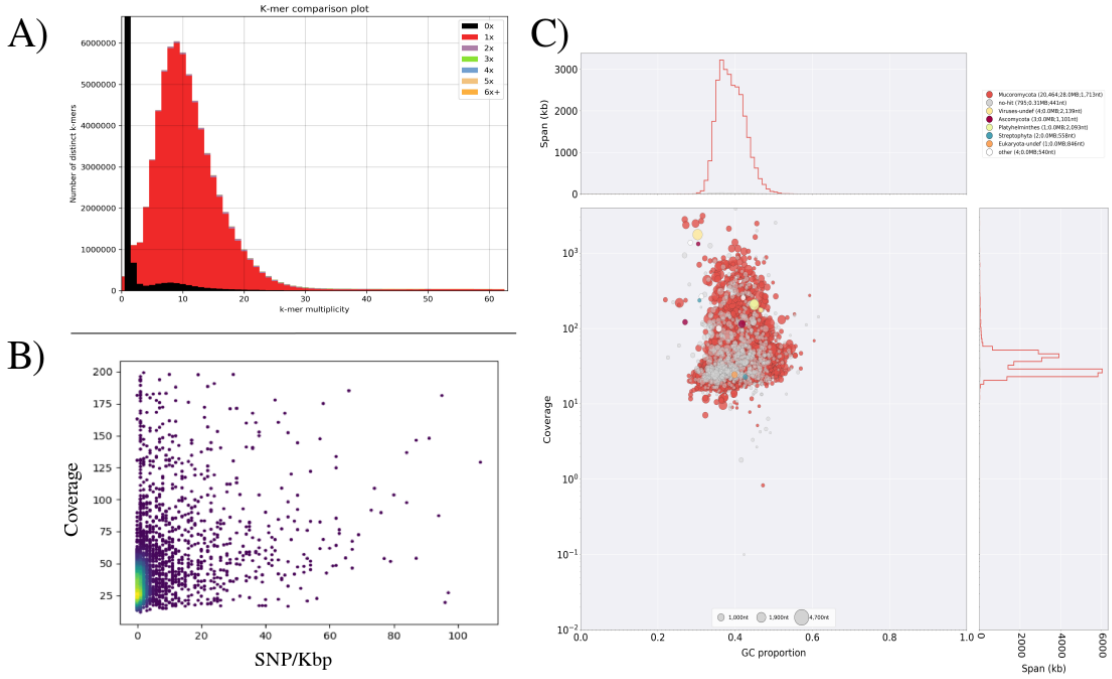

Figure 4. Analysis of *Rhizopus microsporus* B9738
A) KAT *k*-mer plot shows very low genome compaction (black area), suggestive of a haploid genome **B)** Variation versus coverage plot reveals a single main behavior for the genome with regards of its SNP density and coverage **C)** Blobtools analyses shows no sign of widespread contamination that might be inflating the genome.

Our analyses on *Mucor racemosus* B9645 depict a genome with a dual behavior. Distribution of heterozygosity and coverage show two peaks with very low heterozygosity but different coverage (Figure 5). This is further confirmed with the *k*-mer spectrum analysis, which reveals two clear peaks. The genome available in NCBI is 65.5Mbp-long, quite larger than the 45.9Mbp we recovered

in our analyses. The reduction step of Redundans cannot explain this difference, as the assembly size prior to this step is already 46.8Mbp, very close to the final result. Our analyses suggest that contaminating sequences are very minor and do not explain the observed pattern (Figure 5). We hypothesize that *Mucor racemosus* B9645 is a hemidiploid, which presents a portion of its genome in haploid state, and other portion in a highly homozygous diploid state. Due to the low heterozygosity exhibited by this strain, the observed genome architecture might have arisen by either autopolyploidization followed by chromosome loss; or by chromosomal duplications.

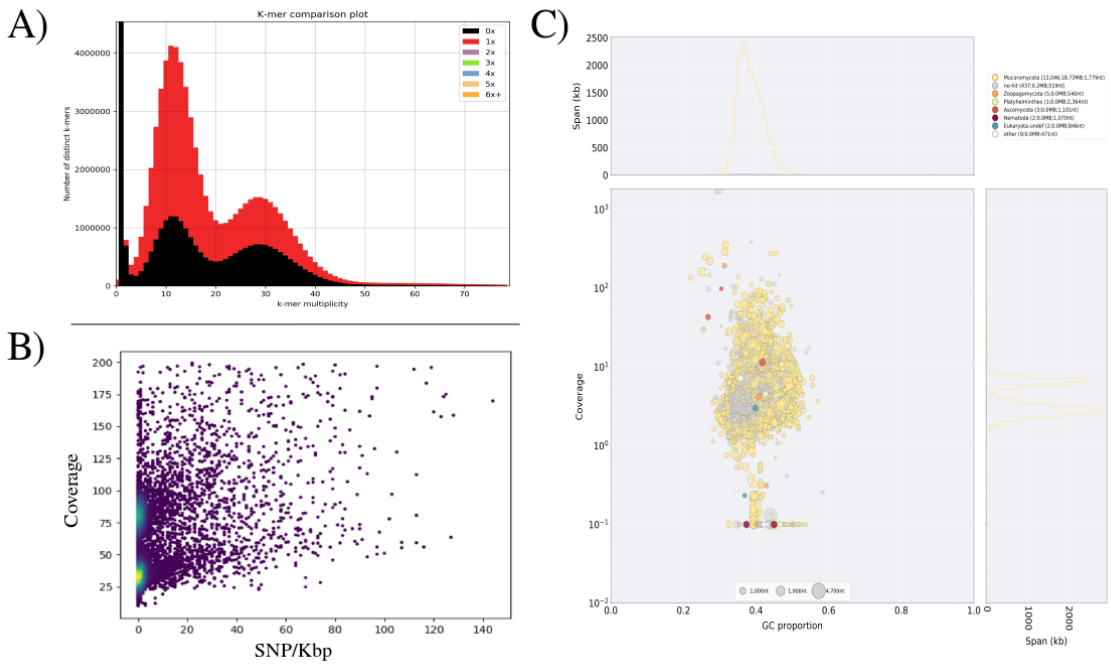

Figure 5. Analysis of *Mucor racemosus* B9645
A) KAT *k*-mer plot shows two peaks with considerable genome compation (black area), suggestive of a diploid genome **B)** Variation versus coverage plot reveals a bimodal behaviour for the genome with regards of its coverage, but both peaks appear with very low SNP density **C)** Blobtools analyses shows no sign of widespread contamination that might be inflating the genome.

The NCBI assembly of *Lichtheimia ramosa* B5399 is 45.6 Mpb long, much larger than the 26.6Mbp we have recovered. Unlike other genomes, our assembly presents a considerable improved quality, going from 3,968 scaffolds and N50 of 33,650 in the NCBI assembly to 861 scaffolds and N50 of 133635 in our own assembly. The relative reduction arises from the use of dipSPAdes, a program made for diploid genome assembly. SPAdes, which is launched during dipSPAdes pipeline, produces an inflated assembly of 67.8Mb. *L. ramosa* presents a heterozygosity level around 3% (Figure 6). All considered, we propose that *L. ramosa* B5399 is a heterozygous diploid, likely resulting from mating between two distantly related strains, and the NCBI assembly is inflated as a consequence of this situation.

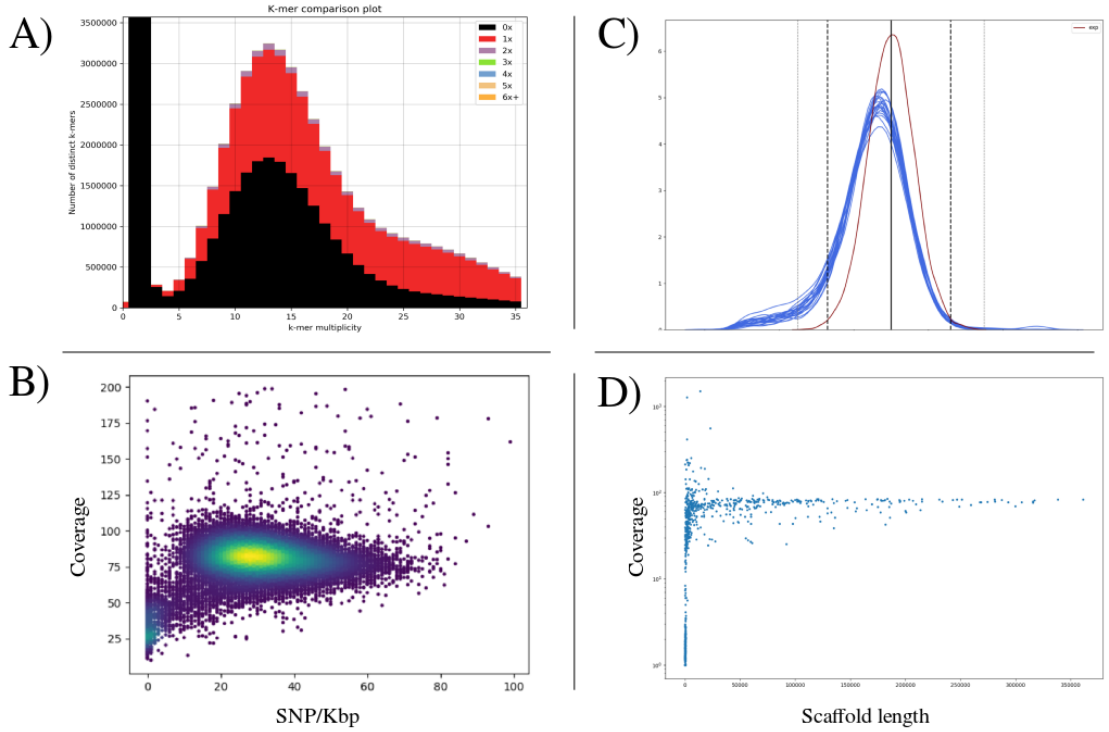

Figure 6. Analysis of *Lichtheimia ramosa* B5399.
A) KAT *k*-mer plot shows one peak with considerable genome compaction (black area) suggestive of a diploid genome **B)** Variation versus coverage plot reveals a unimodal behaviour for the genome with regards of its coverage, presenting a widespread heterozygosity of approximately 3% (maximum density around 30 SNP/Kbp) **C)** Alternative allele frequency shows that all scaffolds in the present a behaviour very similar to the ideal diploid. **D)** Scaffold length plot shows that, with the exception of a group of very low coverage scafflods, all the genome presents a uniform coverage.

5.2.4 Discussion

While the majority of assemblies show no sign of any of the considered biological conditions, we were able to effectively find underlying non-canonical genomic architectures that gone unnoticed in these assemblies. These results suggest that many authors do not take into consideration these kind of genomic accidents, which in turn greatly hampers the results that might be obtained from them. Our results so far are restricted to a relatively narrow clade of Fungi, and these kind of genomic anomalies might be much more common in other groups.

Additionally, there are two strong arguments that suggest that our dataset might not be truly representative of the reality of fungal genomes, even within the taxonomic range we have selected. The first one is the fact that fungal biomass destined to DNA extraction and subsequent sequencing comes from cultures. This implies an important ecological step in which the fungus is forced to grow at optimal speed and in the absence of most stressors. Since aneuploidies, polyploidies and other similar genomic rearrangements are common in the presence of stressors (Anderson *et al.* , 2015; Berman, 2016; Berman *et al.* , 2016; Todd *et al.* , 2017), but seem to be out-competed by euploid cells in optimal conditions (Kumaran *et al.* , 2013; Scott *et al.* , 2017; Zörgö *et al.* , 2013), the act of growing fungi in culture actively selects against chromosomal aberrations. Analogously, many of these chromosomic aberration might exist in nature but are unable to grow on optimal medium, as a component of the fungal "Great Plate Count Anomaly". The advance of environmental sequencing might cast some light in this matter in coming years. The second factor to consider is purely human, as the analyzed datasets have been uploaded by researchers who considered the assemblies were good enough to be uploaded. It is to be expected that many

more low quality assemblies are never deposited and remain in hard drives in laboratories over the world, if not discarded completely.

Even if we consider these biases as non-important, our results recover a significant fraction of publicly available genomes with unorthodox genomic configurations. These have been correlated in many fungal groups with adaptation to novel environments (Kravets *et al.* , 2014; Lenassi *et al.* , 2013; Sinha *et al.* , 2017), resistance to antifungals (Anderson *et al.* , 2017; Harrison *et al.* , 2014), pathogenic capabilities toward both animals (Gerstein *et al.* , 2015; Li *et al.* , 2012; Morrow & Fraser, 2013; Pryszcz, 2014) and plants (Depotter *et al.* , 2016; Garbelotto *et al.* , 2004) and even domestication o industrial strains (Avramova *et al.* , 2018; Borneman *et al.* , 2014; James *et al.* , 2005; Louis *et al.* , 2012; Peter *et al.* , 2018; Walther *et al.* , 2014). Beyond that, contamination in sequencing libraries is a problem that can affect any project, and might cause great problems if left unchecked. Validation of published results goes far beyond the interest of discovering overlooked findings. Comparative genomic studies are limited in their scope and reliability by the quality of assembly and annotation of the genomes, factors that can be greatly compromised by the aforementioned chromosomic phenomena. Comparative studies commonly require the use of flagship genomes that represent a given taxon, often generating a chronology of comparisons with the reference that shapes the perspective on the group. This implies that errors in strategic genomes, such as reference strains or strains in groups with few species, might carry a domino affect over many future works. Long-read sequencing technologies, which are already being widely used for genome assembly projects, hold the promise of providing much more information that could be used to resolve many of these unorthodox genomic architectures. However, these approaches require novel computational approaches to fully employ their potential. Even more, these new technologies still present an extremely high error rate, a highly problematic trait when studying genome heterozygosity.

References

Aguiar, Derek, & Istrail, Sorin. 2013. Haplotype assembly in polyploid genomes and identical by descent shared tracts. *Bioinformatics (Oxford, England)*, **29**(13), i352–60.

Altschul, Stephen F and Gish, Warren and Miller, Webb, Eugene W. Myers, & Lipman, David J. 1990. *BLAST*.

Anderson, Cori A, Roberts, Samantha, Zhang, Huaiying, Kelly, Courtney M, Kendall, Alexxy, Lee, ChangHwan, Gerstenberger, John, Koenig, Aaron B, Kabeche, Ruth, & Gladfelter, Amy S. 2015. Ploidy variation in multinucleate cells changes under stress. *Molecular biology of the cell*, **26**(6), 1129–40.

Anderson, Matthew Z, Saha, Amrita, Haseeb, Abid, & Bennett, Richard J. 2017. A chromosome 4 trisomy contributes to increased fluconazole resistance in a clinical isolate of Candida albicans. *Microbiology (Reading, England)*, **163**(6), 856–865.

Andrews, S. 2010. *FastQC: A quality control tool for high throughput sequence data. Available at: www.bioinformatics.babraham.ac.uk/projects/fastqc/*.

Audano, Peter, & Vannberg, Fredrik. 2014. KAnalyze: A fast versatile pipelined K-mer toolkit. *Bioinformatics*, **30**(14), 2070–2072.

Avramova, Marta, Cibrario, Alice, Peltier, Emilien, Coton, Monika, Coton, Emmanuel, Schacherer, Joseph, Spano, Giuseppe, Capozzi, Vittorio, Blaiotta, Giuseppe, Salin, Franck, Dols-Lafargue, Marguerite, Grbin, Paul, Curtin, Chris, Albertin, Warren, & Masneuf-Pomarede, Isabelle. 2018. Brettanomyces bruxellensis population survey reveals a diploid-triploid complex structured according to substrate of isolation and geographical distribution. *Scientific reports*, **8**(1), 4136.

Benny-Byfield, Simon, & Wendel, Jonathan F. 2014. Doubling down on genomes: Polyploidy and crop plants. *American journal of botany*, **101**(10), 1711–1725.

Berman, Judith. 2016. Ploidy plasticity: a rapid and reversible strategy for adaptation to stress. *FEMS Yeast Research*, **16**(3), fow020.

Berman, Judith, Wertheimer, Noa Blutraich, & Stone, Neil. 2016. Ploidy dynamics and evolvability in fungi. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **371**(20150461).

Bolger, Anthony M, Lohse, Marc, & Usadel, Bjoern. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, **30**(15), 2114–20.

Bonizzoni, Paola, Dondi, Riccardo, Klau, Gunnar W., Pirola, Yuri, Pisanti, Nadia, & Zaccaria, Simone. 2016. On the Minimum Error Correction Problem for Haplotype Assembly in Diploid and Polyploid Genomes. *Journal of Computational Biology*, **23**(9), 718–736.

Borneman, Anthony R, Zeppel, Ryan, Chambers, Paul J, & Curtin, Chris D. 2014. Insights into the Dekkera bruxellensis genomic landscape: comparative genomics reveals variations in ploidy and nutrient utilisation potential amongst wine isolates. *PLoS genetics*, **10**(2), e1004161.

Brudno, Michael, Chapman, Michael, Göttgens, Berthold, Batzoglou, Serafim, Morgenstern, Burkhard, Knowles, S, Bye, JM, Beare, DM, & Dunham, I. 2003. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics*, **4**(1), 66.

Consortium, The Uniprot. 2014. Activities at the Universal Protein Resource (UniProt). *Nucleic acids research*, **42**(Database issue), D191–8.

Corrochano, Luis M, Kuo, Alan, Marcet-Houben, Marina, Polaino, Silvia, Salamov, Asaf, Villalobos-Escobedo, José M, Grimwood, Jane, Álvarez, M Isabel, Avalos, Javier, Bauer, Diane, Benito, Ernesto P, Benoit, Isabelle, Burger, Gertraud, Camino, Lola P, Cánovas, David, Cerdá-Olmedo, Enrique, Cheng, Jan-Fang, Domínguez, Angel, Eliáš, Marek, Eslava, Arturo P, Glaser, Fabian, Gutiérrez, Gabriel, Heitman, Joseph, Henrissat, Bernard, Iturriaga, Enrique A, Lang, B Franz, Lavín, José L, Lee, Soo Chan, Li, Wenjun, Lindquist, Erika, López-García, Sergio, Luque, Eva M, Marcos, Ana T, Martin, Joel, McCluskey, Kevin, Medina, Humberto R, Miralles-Durán, Alejandro, Miyazaki, Atsushi, Muñoz-Torres, Elisa, Oguiza, José A, Ohm, Robin A, Olmedo, María, Orejas, Margarita, Ortiz-Castellanos, Lucila, Pisabarro, Antonio G, Rodríguez-Romero, Julio, Ruiz-Herrera, José, Ruiz-Vázquez, Rosa, Sanz, Catalina, Schackwitz, Wendy, Shahriari, Mahdi, Shelest, Ekaterina, Silva-Franco, Fátima, Soanes, Darren, Syed, Khajamohiddin, Tagua, Víctor G, Talbot, Nicholas J, Thon, Michael R, Tice, Hope, de Vries, Ronald P, Wiebenga, Ad, Yadav, Jagjit S, Braun, Edward L, Baker, Scott E, Garre, Victoriano, Schmutz, Jeremy, Horwitz, Benjamin A, Torres-Martínez, Santiago, Idnurm, Alexander, Herrera-Estrella, Alfredo, Gabaldón, Toni, & Grigoriev, Igor V. 2016. Expansion of Signal Transduction Pathways in Fungi by Extensive Genome Duplication. *Current biology*, **26**(12), 1577–1584.

Depotter, Jasper Rl, Seidl, Michael F, Wood, Thomas A, & Thomma, Bart Phj. 2016. Interspecific hy-

bridization impacts host range and pathogenicity of filamentous microbes. *Current Opinion in Microbiology*, **32**, 7–13.

Gabaldón, Toni, & Alioto, Tyler S. 2016. Whole-Genome Sequencing Recommendations. *Pages 13–42 of: Field Guidelines for Genetic Experimental Designs in High-Throughput Sequencing*.

Garbelotto, Matteo, Gonthier, Paolo, Linzer, Rachel, Nicolotti, Giovanni, & Otrosina, William. 2004. A shift in nuclear state as the result of natural interspecific hybridization between two North American taxa of the basidiomycete complex Heterobasidion. *Fungal genetics and biology : FG & B*, **41**(11), 1046–51.

Gerstein, Aleeza C, & Berman, Judith. 2015. Shift and adapt: the costs and benefits of karyotype variations. *Current opinion in microbiology*, **26**(aug), 130–6.

Gerstein, Aleeza C, Fu, Man Shun, Mukaremera, Liliane, Li, Zhongming, Ormerod, Kate L, Fraser, James A, Berman, Judith, & Nielsen, Kirsten. 2015. Polyploid titan cells produce haploid and aneuploid progeny to promote stress adaptation. *mBio*, **6**(5), e01340–15.

Harrison, Benjamin D, Hashemi, Jordan, Bibi, Maayan, Pulver, Rebecca, Bavli, Danny, Nahmias, Yaakov, Wellington, Melanie, Sapiro, Guillermo, & Berman, Judith. 2014. A tetraploid intermediate precedes aneuploid formation in yeasts exposed to fluconazole. *PLoS biology*, **12**(3), e1001815.

Hirsch, Candice N., & Robin Buell, C. 2013. Tapping the Promise of Genomics in Species with Complex, Nonmodel Genomes. *Annual Review of Plant Biology*, **64**(1), 89–110.

James, Stephen a., Bond, Christopher J., Stratford, Malcolm, & Roberts, Ian N. 2005. Molecular evidence for the existence of natural hybrids in the genus Zygosaccharomyces. *FEMS Yeast Research*, **5**(8), 747–755.

James, Timothy Y., Stenlid, Jan, Olson, Åke, & Johannesson, Hanna. 2008. Evolutionary significance of imbalanced nuclear ratios within heterokaryons of the basidiomycete fungus Heterobasidion parviporum. *Evolution*, **62**(9), 2279–2296.

Kim, Jin-Won, Sang-Gyu Seo, Byung-Ki Jun, Jin-WOn Kim, & Sun-Hyung Kim. 2010. Simple and reliable DNA extraction method for the dark pigmented fungus, Carcospora sojina. *Plant Pathology JOurnal*, **26**(3), 289–292.

Kravets, Anatoliy, Yang, Feng, Bethlendy, Gabor, Sherman, Fred, & Rustchenko, Elena. 2014. Adaptation of Candida albicans to growth on sorbose via monosomy of chromosome 5 accompanied by duplication of another chromosome carrying a gene responsible for sorbose utilization. *FEMS Yeast Research*, **14**(5), 708–713.

Kumaran, Rajaraman, Yang, Shi Yow, & Leu, Jun Yi. 2013. Characterization of Chromosome Stability in Diploid, Polyploid and Hybrid Yeast Cells. *PLoS ONE*, **8**(7).

Laetsch, Dominik R., & Blaxter, Mark L. 2017. BlobTools: Interrogation of genome assemblies. *F1000Research*, **6**, 1287.

Lenassi, Metka, Gostinčar, Cene, Jackman, Shaun, Turk, Martina, Sadowski, Ivan, Nislow, Corey, Jones, Steven, Birol, Inanc, Cimerman, Nina Gunde, & Plemenitaš, Ana. 2013. Whole Genome Duplication and Enrichment of Metal Cation Transporters Revealed by De Novo Genome Sequencing of Extremely Halotolerant Black Yeast Hortaea werneckii. *PLoS ONE*, **8**(8), e71328.

Li, Heng, & Durbin, Richard. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**(14), 1754–60.

Li, Heng, Handsaker, Bob, Wysoker, Alec, Fennell, Tim, Ruan, Jue, Homer, Nils, Marth, Gabor, Abecasis, Goncalo, Durbin, Richard, & 1000 Genome Project Data Processing Subgroup. 2009a. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**(16), 2078–9.

Li, Wenjun, Averette, Anna Floyd, Desnos-Ollivier, Marie, Ni, Min, Dromer, Françoise, & Heitman, Joseph. 2012. Genetic Diversity and Genomic Plasticity of Cryptococcus neoformans AD Hybrid Strains. *G3: Genes, Genomes, Genetics*, **2**(1), 83–97.

Louis, V. L., Despons, L., Friedrich, A., Martin, T., Durrens, P., Casaregola, S., Neuveglise, C., Fairhead, C., Marck, C., Cruz, J. a., Straub, M.-L., Kugler, V., Sacerdot, C., Uzunov, Z., Thierry, A., Weiss, S., Bleykasten, C., De Montigny, J., Jacques, N., Jung, P., Lemaire, M., Mallet, S., Morel, G., Richard, G.-F., Sarkar, A., Savel, G., Schacherer, J., Seret, M.-L., Talla, E., Samson, G., Jubin, C., Poulain, J., Vacherie, B., Barbe, V., Pelletier, E., Sherman, D. J., Westhof, E., Weissenbach, J., Baret, P. V., Wincker, P., Gaillardin, C., Dujon, B., & Souciet, J.-L. 2012. Pichia sorbitophila, an Interspecies Yeast Hybrid, Reveals Early Steps of Genome Resolution After Polyploidization. *G3: Genes, Genomes, Genetics*, **2**(2), 299–311.

Ma, Li-Jun, Ibrahim, Ashraf S, Skory, Christopher, Grabherr, Manfred G, Burger, Gertraud, Butler, Margi, Elias, Marek, Idnurm, Alexander, Lang, B Franz, Sone, Teruo, Abe, Ayumi, Calvo, Sarah E, Corrochano, Luis M, Engels, Reinhard, Fu, Jianmin, Hansberg, Wilhelm, Kim, Jung-Mi, Kodira, Chinnappa D, Koehrsen, Michael J, Liu, Bo, Miranda-Saavedra, Diego, O'Leary, Sinead, Ortiz-Castellanos, Lucila, Poulter, Russell, Rodriguez-Romero, Julio, Ruiz-Herrera, José, Shen, Yao-Qing, Zeng, Qiandong, Galagan, James, Birren, Bruce W, Cuomo, Christina A, & Wickes, Brian L. 2009. Genomic analysis of the basal lineage fungus Rhizopus oryzae reveals a whole-genome duplication. *PLoS genetics*, **5**(7), e1000549.

Maheshwari, Ramesh. 2005. Nuclear behavior in fungal hyphae. *FEMS Microbiology Letters*, **249**, 7–14.

Mamun, Abdullah-Al, Pal, Soumitra, & Rajasekaran, Sanguthevar. 2016. KCMBT: a k -mer Counter based on Multiple Burst Trees. *Bioinformatics*, **32**(18), 2783–2790.

Mapleson, Daniel, Accinelli, Gonzalo Garcia, Kettleborough, George, Wright, Jonathan, Clavijo, Bernardo J., Garcia Accinelli, Gonzalo, Kettleborough, George, Wright, Jonathan, & Clavijo, Bernardo J. 2016. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*, **9**(4), btw663.

Marcais, G, & Kingsford, C. 2012. Jellyfish : A fast k-mer counter. *Tutorialis e Manuais*, 1–8.

McKenna, Aaron, Hanna, Matthew, Banks, Eric, Sivachenko, Andrey, Cibulskis, Kristian, Kernytsky,

Andrew, Garimella, Kiran, Altshuler, David, Gabriel, Stacey, Daly, Mark, & DePristo, Mark A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, **20**(9), 1297–303.

Morrow, Carl a., & Fraser, James a. 2013. Ploidy variation as an adaptive mechanism in human pathogenic fungi. *Seminars in Cell and Developmental Biology*, **24**(4), 339–346.

National Centre for Biotechnology Information. 2015. SRA: Sequence Read Archive. *NCBI Handout Series*, 4.

Neale, David B, Wegrzyn, Jill L, Stevens, Kristian a, Zimin, Aleksey V, Puiu, Daniela, Crepeau, Marc W, Cardeno, Charis, Koriabine, Maxim, Holtz-Morris, Ann E, Liechty, John D, Martínez-García, Pedro J, Vasquez-Gross, Hans a, Lin, Brian Y, Zieve, Jacob J, Dougherty, William M, Fuentes-Soriano, Sara, Wu, Le-Shin, Gilbert, Don, Marçais, Guillaume, Roberts, Michael, Holt, Carson, Yandell, Mark, Davis, John M, Smith, Katherine E, Dean, Jeffrey Fd, Lorenz, W Walter, Whetten, Ross W, Sederoff, Ronald, Wheeler, Nicholas, McGuire, Patrick E, Main, Doreen, Loopstra, Carol a, Mockaitis, Keithanne, Dejong, Pieter J, Yorke, James a, Salzberg, Steven L, & Langley, Charles H. 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome biology*, **15**(3), R59.

Otto, Sarah P. 2007. The Evolutionary Consequences of Polyploidy. *Cell*, **131**(3), 452–462.

Peter, Jackson, De Chiara, Matteo, Friedrich, Anne, Yue, Jia-Xing, Pflieger, David, Bergström, Anders, Sigwalt, Anastasie, Barre, Benjamin, Freel, Kelle, Llored, Agnès, Cruaud, Corinne, Labadie, Karine, Aury, Jean-Marc, Istace, Benjamin, Lebrigand, Kevin, Barbry, Pascal, Engelen, Stefan, Lemainque, Arnaud, Wincker, Patrick, Liti, Gianni, & Schacherer, Joseph. 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates Species-wide genetic and phenotypic diversity. *Nature*, **556**.

Pryszcz, Leszek P, & Gabaldón, Toni. 2016. Redundans : an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Research*, **8**(44), 1–16.

Pryszcz, Leszek Piotr. 2014. *Comparative genomics to unravel virulence mechanisms in fungal human pathogens*. Ph.D. thesis.

Safonova, Yana, Bankevich, Anton, & Pevzner, Pavel A. 2015. dipSPAdes: Assembler for Highly Polymorphic Diploid Genomes. *Journal of computational biology : a journal of computational molecular cell biology*, **22**(6), 528–45.

Salman-Minkov, Ayelet, Sabath, Niv, & Mayrose, Itay. 2016. Whole-genome duplication as a key factor in crop domestication. *Nature Plants*, **2**(8), 16115.

Schoenfelder, Kevin P, & Fox, Donald T. 2015. The expanding implications of polyploidy. *The Journal of cell biology*, **209**(4), 485–91.

Scott, Amber L, Richmond, Phillip A, Dowell, Robin D, & Selmecki, Anna M. 2017. The Influence of Polyploidy on the Evolution of Yeast Grown in a Sub-Optimal Carbon Source. *Molecular biology and evolution*, **34**(10), 2690–2703.

Session, Adam M., Uno, Yoshinobu, Kwon, Taejoon, Chapman, Jarrod A., Toyoda, Atsushi, Takahashi, Shuji, Fukui, Akimasa, Hikosaka, Akira, Suzuki, Atsushi, Kondo, Mariko, van Heeringen, Simon J., Quigley, Ian, Heinz, Sven, Ogino, Hajime, Ochi, Haruki, Hellsten, Uffe, Lyons, Jessica B., Simakov, Oleg, Putnam, Nicholas, Stites, Jonathan, Kuroki, Yoko, Tanaka, Toshiaki, Michiue, Tatsuo, Watanabe, Minoru, Bogdanovic, Ozren, Lister, Ryan, Georgiou, Georgios, Paranjpe, Sarita S., van Kruijsbergen, Ila, Shu, Shengquiang, Carlson, Joseph, Kinoshita, Tsutomu, Ohta, Yuko, Mawaribuchi, Shuuji, Jenkins, Jerry, Grimwood, Jane, Schmutz, Jeremy, Mitros, Therese, Mozaffari, Sahar V., Suzuki, Yutaka, Haramoto, Yoshikazu, Yamamoto, Takamasa S., Takagi, Chiyo, Heald, Rebecca, Miller, Kelly, Haudenschild, Christian, Kitzman, Jacob, Nakayama, Takuya, Izutsu, Yumi, Robert, Jacques, Fortriede, Joshua, Burns, Kevin, Lotay, Vaneet, Karimi, Kamran, Yasuoka, Yuuri, Dichmann, Darwin S., Flajnik, Martin F., Houston, Douglas W., Shendure, Jay, DuPasquier, Louis, Vize, Peter D., Zorn, Aaron M., Ito, Michihiko, Marcotte, Edward M., Wallingford, John B., Ito, Yuzuru, Asashima, Makoto, Ueno, Naoto, Matsuda, Yoichi, Veenstra, Gert Jan C., Fujiyama, Asao, Harland, Richard M., Taira, Masanori, & Rokhsar, Daniel S. 2016a. Genome evolution in the allotetraploid frog Xenopus laevis. *Nature*, **538**(7625), 336–343.

Simpson, Jared T., & Pop, Mihai. 2015. The Theory and Practice of Genome Sequence Assembly. *Annual Review of Genomics and Human Genetics*, **16**(1), 153–172.

Sinha, Sunita, Flibotte, Stephane, Niera, Mauricio, Formby, Sean, Plemenitaš, Ana, Gunde Cimerman, Nina, Lenassi, Metka, Gostinčar, Cene, Stajich, Jason E., & Nislow, Corey. 2017. Insight into the Recent Genome Duplication of the Halophilic Yeast Hortaea werneckii: Combining an Improved Genome with Gene Expression and Chromatin Structure. *G3: Genes, Genomes, Genetics*.

Sreenivasaprasad, Surapareddy. 2000. Isolation of Fungal Nucleic Acids. *Pages 37–45 of: The Nucleic Acid Protocols Handbook*. New Jersey: Humana Press.

Strom, Noah B, & Bushley, Kathryn E. 2016. Two genomes are better than one: history, genetics, and biotechnological applications of fungal heterokaryons. *Fungal biology and biotechnology*, **3**, 4.

Strope, Pooja K, Skelly, Daniel A, Kozmin, Stanislav G, Mahadevan, Gayathri, Stone, Eric A, Magwene, Paul M, Dietrich, Fred S, & McCusker, John H. 2015a. The 100-genomes strains, an S. cerevisiae resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome research*, **25**(5), 762–74.

Todd, Robert T, Forche, Anja, & Selmecki, Anna. 2017. Ploidy Variation in Fungi: Polyploidy, Aneuploidy, and Genome Evolution. *Microbiology spectrum*, **5**(4).

Torres, Eduardo M, Williams, Bret R, & Amon, Angelika. 2008. Aneuploidy: cells losing their balance. *Genetics*, **179**(2), 737–46.

Wajid, Bilal, & Serpedin, Erchin. 2012. Review of General Algorithmic Features for Genome Assemblers for Next Generation Sequencers. *Genomics, Proteomics & Bioinformatics*, **10**, 58–73.

Wajid, Bilal, Sohail, Muhammad U, Ekti, Ali R, & Serpedin, Erchin. 2016. The A, C, G, and T of Genome Assembly. *BioMed research international*, **2016**, 6329217.

Walther, Andrea, Hesselbart, Ana, & Wendland, Jürgen. 2014. Genome Sequence of Saccharomyces carlsbergensis, the World's First Pure Culture Lager Yeast. *G3: Genes, Genomes, Genetics*, **4**(5), 1–11.

Weiss, Clemens L., Pais, Marina, Cano, Liliana M., Kamoun, Sophien, & Burbano, Hernán A. 2017. nQuire: A Statistical Framework For Ploidy Estimation Using Next Generation Sequencing. *bioRxiv*, may, 143537.

Weissenbach, Jean. 2016. The rise of genomics. *Comptes Rendus Biologies*, **339**(7), 231–239.

Wijayawardena, Bhagya K, Minchella, Dennis J, & DeWoody, J Andrew. 2013. Hosts, parasites, and horizontal gene transfer. *Trends in parasitology*, **29**(7), 329–38.

Wilkening, Stefan, Tekkedil, Manu M., Lin, Gen, Fritsch, Emilie S., Wei, Wu, Gagneur, Julien, Lazinski, David W., Camilli, Andrew, & Steinmetz, Lars M. 2013. Genotyping 1000 yeast strains by next-generation sequencing. *BMC Genomics*, **14**(1).

Zhu, Yuan O., Sherlock, Gavin, & Petrov, Dmitri A. 2016. Whole Genome Analysis of 132 Clinical Saccharomyces cerevisiae Strains Reveals Extensive Ploidy Variation. *G3: Genes, Genomes, Genetics*, **6**(8).

Zörgö, Enikö, Chwialkowska, Karolina, Gjuvsland, Arne B., Garré, Elena, Sunnerhagen, Per, Liti, Gianni, Blomberg, Anders, Omholt, Stig W., & Warringer, Jonas. 2013. Ancient Evolutionary Trade-Offs between Yeast Ploidy States. *PLoS Genetics*, **9**(3).

Chapter 6

Phylogenomic analyses within the 1KFG project

Marina Marcet-Houben*, Miguel A Naranjo-Ortiz*, Joseph Spatafora, Toni Gabaldón

*These authors contributed equally

Abstract

The 1KFG (1000 Fungal Genomes) project is an international project that aims to sequence representative species from all branches of the fungal tree of life. As part of this consortium, we have performed a diverse array of phylogenomic analyses that aim to resolve the phylogeny of two conflicting lineages, the Taphrinomycetes and the Wallemiomycetes; as well as to map gene duplication frequencies across the fungal tree of life.

6.1 Introduction

As reviewed in chapter 3, ancient polyploidy events are, so far, anecdotal in Fungi. This situation greatly contrasts with that of plants and animals, for which many events have been identified. It has been suggested that of this difference may be due to the difficulty of recognizing such events in fungi (Campbell *et al.* , 2016). In order to investigate the existence of previously undescribed paleopolyploidization events in Fungi we have constructed a collection of 32 new phylomes -i.e. genome-wide collections of gene evolutionary histories- that collectively include 350 species across the fungal tree of life. We have analyzed this dataset in order to find nodes within the fungal tree of life with high (>0.25) gene duplication frequencies (i.e. average number of duplications per gene in a given lineage). Our results suggest the existence of several potential paleopolyploidization events across Fungi. Duplication peaks are extremely frequent in Fungi, although the distribution of such events is extremely uneven. Additionally, we have used this data to construct a phylogenomic dataset to address the phylogenetic position of three controversial fungal clades: Wallemiomycetes, Archaeorhizomycetes and the relative positions of Blastocladiomycota and Chytridiomycota.

6.2 Material and methods

Phylome reconstruction: For each phylome, a set of species is selected that covers a given taxonomic range. In our case numerous phylomes were reconstructed so that jointly they cover all of the species included so far in the 1KFG (table 1). Phylomes are mostly centered around a single taxonomic class though usually a few members of other classes are included to serve as outgroups. Once the species set is determined, a seed species is chosen, this species serves as the starting point for the homology search analysis. The choice of seed also implies that 100% of the genes of this species will be included in the phylome as long as they have homologs in the rest of the species set. For the additional species only those genes that have homologs in the seed species will be covered. Multiple phylomes can be built using the same species set but using different seed species. Once the set of species is chosen the phylome is reconstructed using an established pipeline which has been earlier explained in detail (Huerta-Cepas *et al.* , 2011). Briefly, for each gene encoded in the seed genome a homology search is performed against the chosen species set using blast (Altschul, Stephen F and Gish, Warren and Miller, Webb & Lipman, 1990). Results are filtered so that only hits with an e-value below 1e-05 and an overlap of at least 50% over the query sequence are chosen. The 150 top hits are chosen for the tree reconstruction step. A multiple sequence alignment is build using the following method: first six multiple alignments are reconstructed, three in forward and three in reverse (Landan & Graur, 2007) using three different programs (Muscle v3.8 (Edgar, 2004), MAFFT v6.861 (Katoh *et al.* , 2005) and Kalign v2.04 (Lassmann & Sonnhammer, 2005). A consensus alignment is then obtained using M-coffee version 10.00.r1607 (Wallace *et al.* , 2006). This alignment is trimmed using a consistency score and a gap threshold using trimAl v1.4 (Capella-Gutiérrez *et al.* , 2009). This cleaned alignment is then used to reconstruct maximum likelihood phylogenetic trees. First the best evolutionary model for each protein is chosen by calculating the likelihood of each of the seven possible models using a neighbor joining tree as calculated by bioNJ (Gascuel, 1997). The best model according to the AIC criterion is then chosen to build a maximum likelihood tree using PhyML v3.0 (Le & Gascuel, 2008). Branch support is calculated using the aLRT algorithm.

Species tree reconstruction: For each phylome a species tree is reconstructed. This tree is automatically reconstructed using a gene tree parsimony approach that minimizes the number of inferred duplications as implemented in duptree v1.48 (Wehe *et al.* , 2008), and using the entire phylome. When the species tree does not match the expected species tree, coherent with known phylogenetic relationships, then an alternative topology is used for the analyses. For some specific analyses a concatenated species tree was reconstructed as follows: Species specific duplications were deleted from the phylome trees and one sequence at random was kept. Then trees that did not contain any duplication were selected. Those that had homologs in a given number of species were chosen to perform the species tree reconstruction. Alignments of the chosen trees were filtered

to remove species specific duplications if necessary and were then concatenated. RaxML (v8.2.4) (Stamatakis, 2014) was used to reconstruct the species trees using the PROTGAMMALG model.

Phylome analysis and duplication rates: Each phylome was analyzed using ETE v3.1 (Huerta-Cepas *et al.* , 2016). Orthology and paralogy relationships were established between the seed protein and its homologs according to a species overlap algorithm (Gabaldón, 2008). This algorithm assumes a duplication is present at a given node of a gene tree when an overlap exists between the species at the two sides of the node. Each duplication is then mapped onto the species tree assuming that it occurred at the common ancestor of all the species included in the node (Huerta-Cepas & Gabaldon, 2011). Only duplications found in the lineage leading from the root to the seed protein are considered. Duplications for which the node support is below 0.95 are excluded. The number of duplications that map to a given node is then divided by the number of gene trees that include that node. The result of this calculation is the duplication ratio (DR) which can serve as indicator to the presence of a polyploidization event.

Topology testing: ETE v3.1 was used to search phylomes for specific topologies. First the alternative topologies were established and then the trees were scanned to search for those topologies. For each tree only the clade containing the seed protein was considered. For a node to be used for topology testing it and its parent node have to be highly supported (aLRT >= 0.95). For the Blastocladiomycota / Chytridiomycota question different thresholds were used to analyze trees. 1.- We search for the first group that does not belong to the same group as the seed and see which species are found there. 2.- Same as 1 but we first remove species specific duplications and then we only use trees without duplications. 3.- We only use trees where the groups considered, except for outgroups, are monophyletic: Blastocladiomycota, Chytridiomycota, other fungi. In the large phylome there was no tree fulfilling this last scenario and we applied it only on the smaller phylomes. This last analysis, while drastically reducing the dataset, is likely the most reliable.

CONSEL: Alternative phylogenetic trees were reconstructed using ETE v3.1. Likelihoods per site were calculated using phyml for each alternative topology using the concatenated species tree used to reconstruct the initial tree. Then CONSEL was used to test whether any of the topologies reconstructed could be discarded.

Explorer: We implemented a sub-sampling based algorithm that calculates duplication ratios for a large set of species. Given the complete set of species, 1000 random proteins are selected for each species. Then, phylogenetic trees are reconstructed using a fast method. First a homology search is performed against the complete species set. Homologs are chosen using the same parameters as before. Alignments are then build using MAFFT v6.861 (Katoh & Standley, 2013) and Fasttree 2.1.8 SSE3 (Price *et al.* , 2010) is used to reconstruct a phylogenetic tree using the WAG model. Once all trees are reconstructed, duplication rates are calculated as explained above. We have also calculated an additional duplication measure (DR2) that omits counting multiple duplications at the same node for one protein. This measure will point to the number of trees that have a duplication at a certain node without being affected by multiple duplications.

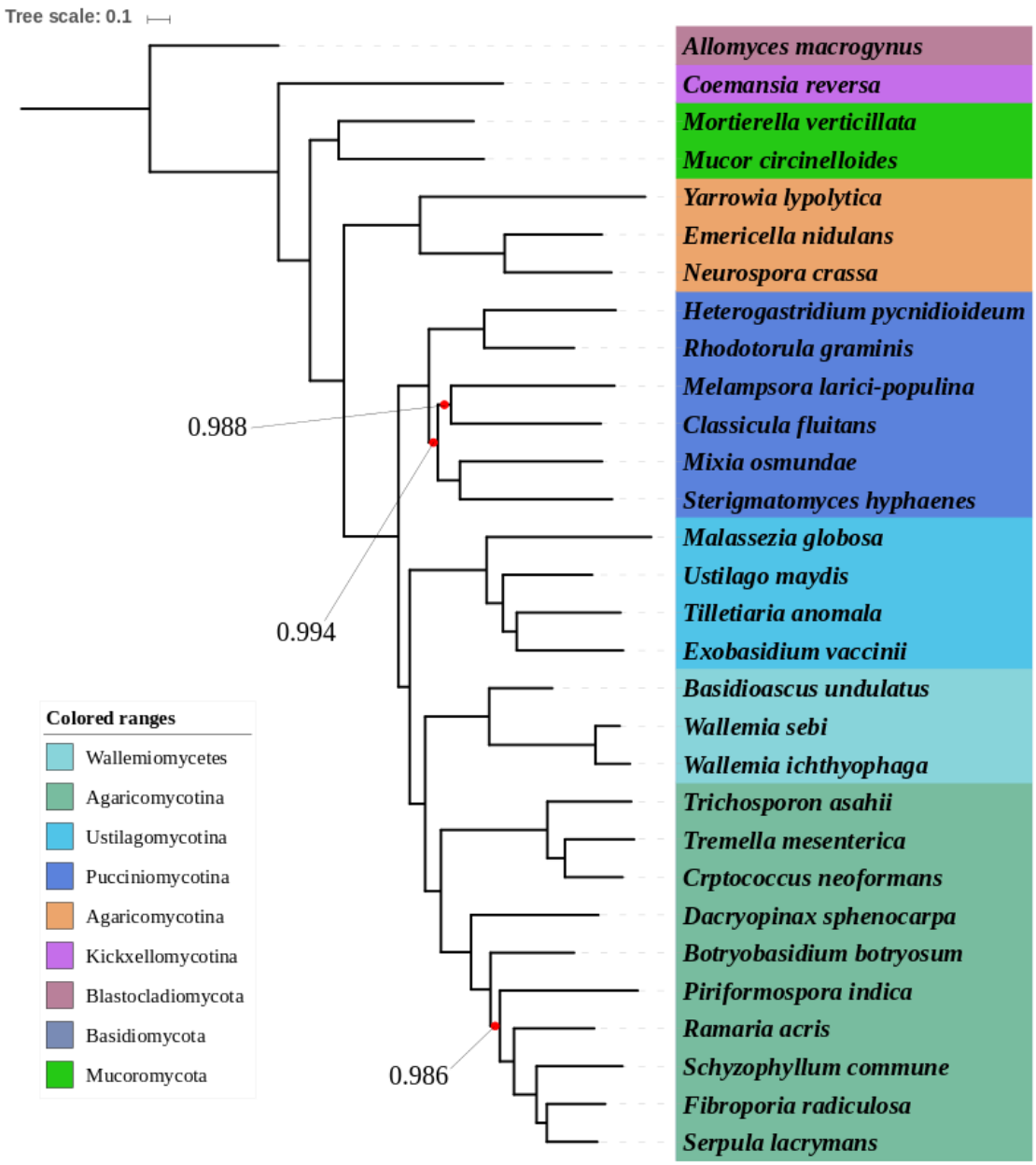

Figure 1. Phylogeny of Wallemiomycetes

Phylogenetic placement of the Wallemiomycetes, based of phylogenomic reconstruction using the phylome of *Wallemia sebi*. Nodes with a bootstrap support below 1 are indicated by a red dot and its bootstrap support value shown. Tree visualization has been performed using the interactive Tree of Life server (iTOL) (Letunic & Bork, 2016).

6.3 Results

6.3.1 Phylogenetic placement of Wallemiomycetes

Wallemiomycetes is a class with three genera of highly extremotolerant fungi of uncertain phylogenetic position within the Basidiomycota. The class includes the genera *Wallemia*, *Basidioascus* and *Geminibasidium*, for which genomic information is currently available for only the first two. The phylogenetic position of Wallemiomycetes is controverted, with some phylogenetic studies recovering the group as sister to *bona fide* Agaricomycotina, often including it as a class within the subphylum (Mishra *et al.* , 2018; Nguyen *et al.* , 2015a; Padamsee *et al.* , 2012; Zajc *et al.* , 2013); while others recover the group in other positions within the Basidiomycota and propose the elevation to Subphylum Wallemiomycetes (Bauer *et al.* , 2015; Matheny *et al.* , 2006; Q.-M. Wang *et al.* , 2014). The conflicting positioning seems to stem from the long branches presented on *Wallemia*, and thus should be possible to solve it with the incorporation of data from *Basidioascus* or *Geminibasidium*.

We selected 259 genes that had one-to-one orthologs in the 30 species included in the *Wallemia sebi* phylome (760) and concatenated their trimmed alignments. The final alignment containing 165,059 amino acid positions was used to reconstruct a maximum likelihood species tree with PhyML v3.0 (Guindon *et al.* , 2010) using LG as a model, and estimating four rate categories and the fraction of invariant sites from the alignments. Supports for the topology were computed using 100 bootstrap alignment replicates. The final tree (Fig. 1) recovers with high support Wallemiomycotina, here represented by *Wallemia sebi*, *Wallemia ichthyophaga* and *Basidioascus undulatus*; as sister group to Agaricomycotina.

6.3.2 Phylogenetic placement of Archaeorhizomycetes

Archaeorhizomycetes is a recently described class of root endophytes that was first identified through environmental sequences. Phylogenetic analyses place the clade within Taphrinomycotina. However, all lineages in the group present very small genomes and several (I.e *Schizosacharomyces*, *Pneumocystis*, *Saitoella*) (Kuramae *et al.* , 2006; Liu *et al.* , 2008; Schoch *et al.* , 2009a; Sugiyama *et al.* , 2006) also seem to have suffered accelerated evolution. As a result, the phylogenetic relationships between the main classes of Taphrinomycotina are controversial, specially with regards of the placement of *Saitoella complicata*.

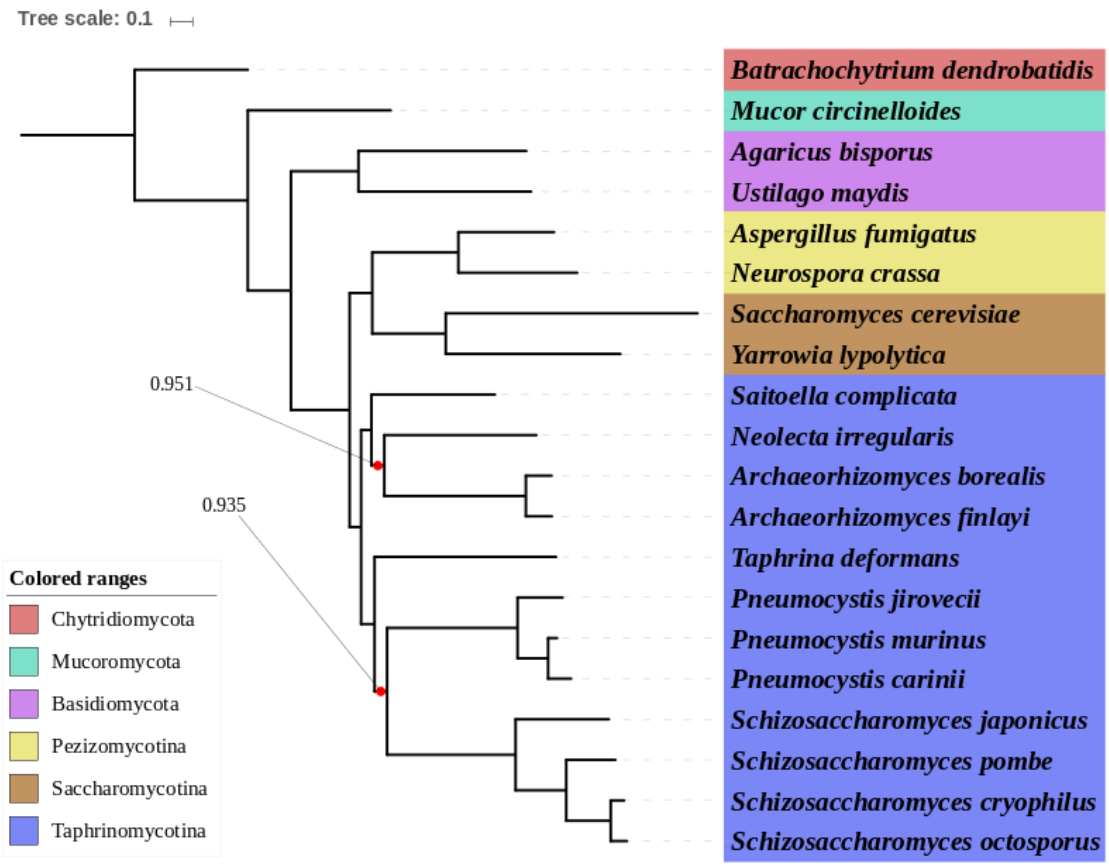

Figure 2. Phylogeny of Taphrinomycotina

Phylogenetic relationships among the different lineages of Taphrinomycotina, based of phylogenomic reconstruction using the phylome of *Schizosaccharomyces pombe*. Nodes with a bootstrap support below 1 are indicated by a red dot and its bootstrap support value shown. Tree visualization has been performed using the interactive Tree of Life server (iTOL) (Letunic & Bork, 2016).

We selected 186 genes that had one-to-one orthologs in at least 19 of the 20 species included in *Schizosaccharomyces pombe* phylome 767 and concatenated their trimmed alignments, as reconstructed in the phylome. The final alignment containing 116,939 amino acid positions was used to reconstruct the maximum likelihood species tree as described above. Supports for the topology were computed using 100 bootstrap alignment replicates. The final tree (Fig. 2) recovers a highly-supported clade containing *Saitoella complicata*, *Neolecta irregularis* and *Archaeorhizomyces*, thus indicating that the later is a *bona fide* member of the Taphrinomycotina. Taphrinomycotina appears as a monophyletic clade with high support, and is placed as the earliest diverging group within Ascomycota. *Saitoella* appears as sister to *Archaeorhizomyces* + *Neolecta* (Bootstrap=0.951), forming the first split in the Taphrinomycotina. Sisterhood of *Archaeorhizomyces* and *Neolecta* is in agreement with previous studies (Menkis *et al.* , 2014). The second split we recover for the Taphrinomycotina contains *Taphrina deformans* as sister to a clade including the different species of *Pneumocystis* and *Schizosaccharomyces* (Bootstrap=0.931). Sisterhood of *Schizosaccharomyces* and *Pneumocstis* has been recovered by other authors (Liu *et al.* , 2008; Menkis *et al.* , 2014; Schoch *et al.* , 2009a). Our results indicate that the class Taphrinomycetes is currently paraphyletic, and thus the taxonomy of *Saitoella* should be reviewed accordingly. However, the relationships between the different classes within Taphrinomycotina have not be solved confidently.

6.3.3 Phylogenetic inconsistency between Blastocladiomycota and Chytridiomycota

Two of the most ancestral fungal groups are Blastocladiomycota and Chytridiomycota. Their relative positions in the fungal tree of life based mainly on gene markers is not yet resolved (James *et al.* , 2014). We used phylomes in order to address this question by comparing support to three different hypotheses. Given a fungal tree of life where the first species that diverged were the Microsporidia and the Cryptomycota, the two following groups are the Blastocladiomycota and the Chytridiomycota. The three possible topologies are the following: either Blastocladiomycota (topology A) or Chytridiomycota (topology B) diverged first or they have a common ancestor and form a monophyletic group (topology C). We reconstructed two initial phylomes that contained 30 species and in which each has a seed species belonging to one of the two groups (phylomes 416 – *Catenaria anguillulae* (Blastocladiomycota) and 417 – *Batrachochytrium dendrobatidis* (Chytridiomycota)). Then, for each phylome, we performed a topological analysis to assess how many trees supported each of the three alternative topologies. The results of this analysis were inconclusive. Topology A was slightly better supported but the difference was very small. We applied stricter filters (see methods) but no better pattern emerged. In order to obtain a clearer picture, we reconstructed two smaller phylomes that included less additional species and used the same seed species (see phylome 418 and 419). The patterns here are clearer when the strictest filters are applied (support >= 0.95 of the nodes of interest, monophyly of the different taxonomic groups), with roughly 50% of the tree supporting topology A while the two alternative topologies are equally supported. This pattern is typical in cases of incomplete lineage sorting. We performed the same analysis dividing the trees in three bins according to their alignment length. We noticed that, the longer the alignment, the more support there is for topology A though the differences are not very big. Still, using such strict parameters greatly reduces our dataset and we are only analyzing about 500 trees of the two phylomes. We reconstructed a phylogenetic species tree based on 284 genes that were single copy in at least 27 of the 30 species included in the first set of phylomes (see methods). The resultant topology was congruent with the previous analysis and showed a basal position of Blastocladiomycota. We then used ETE v3.1 to obtain two derived trees that presented the alternative topologies described above. Then we used CONSEL to see whether we could discard any of the three hypotheses. According to the AU test implemented in CONSEL we can discard the hypothesis that Blastocladiomycota and Chytridiomycota group together (topology C), but we cannot discard topology B. So, while the topology testing above for the smaller phylomes seems to point to a basal position of Blastocladiomycota, additional species diminish the signal and makes us unable to distinguish which is the correct topology. This could be partially due to the low number of species in these groups or because of the complicated evolutionary history found in early diverging fungi.

6.3.4 Duplication frequency across the fungal tree of life

Global trends To assess how gene duplications are distributed over the fungal tree of life we reconstructed 291,885 fast trees (See methodology) from random subsets of 1,000 genes from each of the 352 representative species comprising the 1KFG pilot phase, as well as 32 complete phylomes comprising 198,374 accurate trees and covering 350 of the species in the set (see Materials and Methods). Using a species-overlap coupled to a relative dating strategy (Huerta-Cepas & Gabaldón, 2011), we detected and dated gene duplications to the different nodes of the 1KFG species tree. We set an arbitrary threshold of 0.25 duplications per gene to define 99 (14.32%) nodes as duplication peaks (i.e. nodes with a high duplication ratio). Our results (Figure 1) show an intriguing pattern in the distribution of duplication ratios (DR) peaks throughout the fungal tree of life. As seen in figure 1, duplication peaks seem to be common in all fungal lineages with the glaring exception of the Pezizomycotina. Ustilaginomycotina and Taphrinomycotina also show very few of these peaks although the number of species represented in these clades is limited (11 and 12 species, respectively). The same cannot be said about Pezizomycotina, which with 153 species is the largest and best represented group in our dataset. In the following sections we describe relevant findings in the different fungal clades.

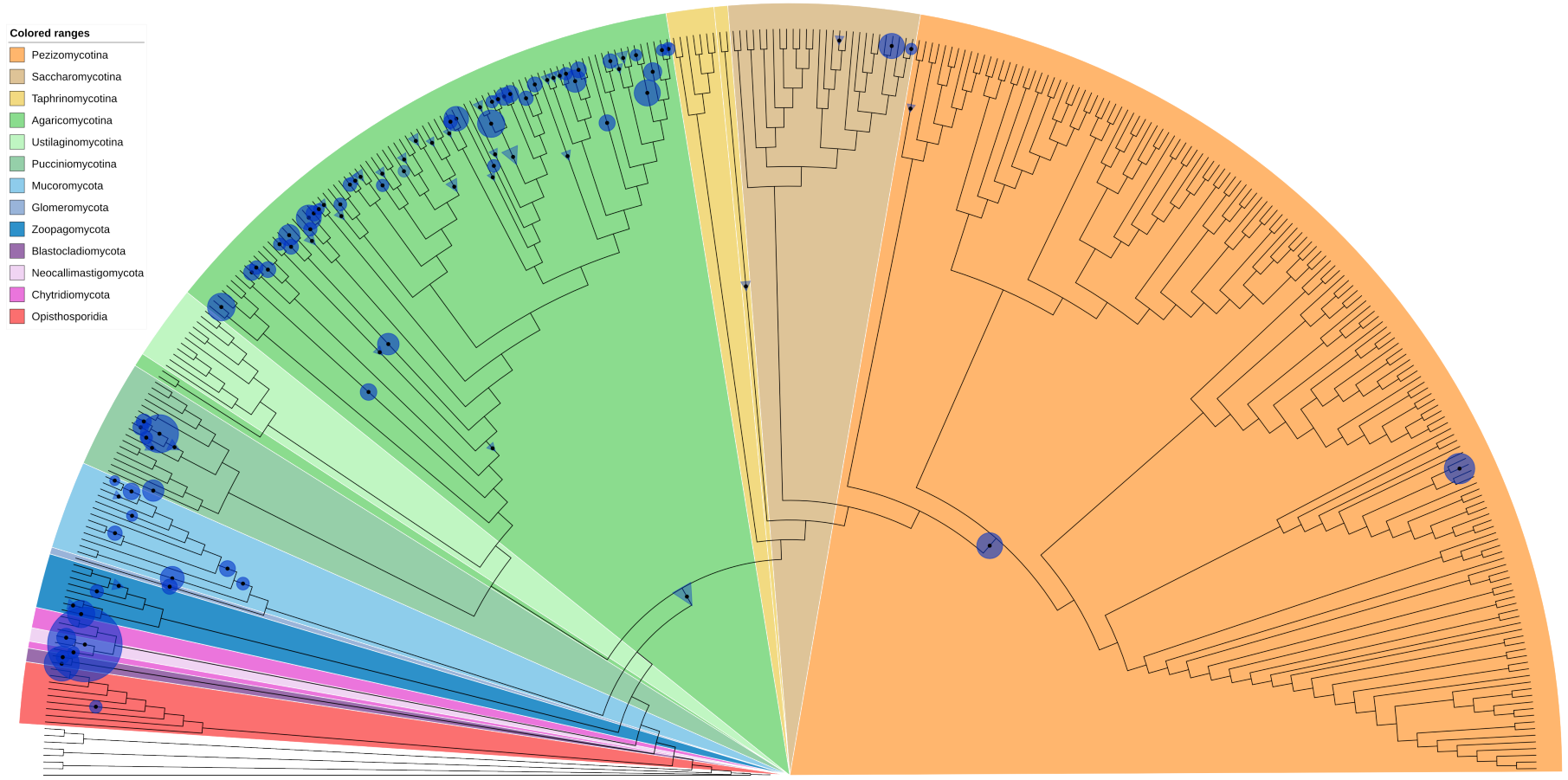

Figure 3. Duplication rate across the fungal tree of life

Phylogenetic tree covering all represented species in the collection of phylomes. For the sake of simplicity, individual leaf names have been omitted. Nodes with a duplication rate (DR) above 0.25 according to explorer calculations are marked with a symbol whose diameter is proportional to the value. Nodes with an corrected duplication rate (DR2) below 0.25 are marked with a triangle, otherwise they are marked with a circle. Tree visualization has been performed using the interactive Tree of Life server (iTOL) (Letunic & Bork, 2016).

Early diverging lineages Duplication peaks appear to be very common among Non-Dikarya fungi (Figure 4), even though the representation of these groups in our dataset is rather poor. Despite the extreme genomic reduction in Microsporidia, we observe a considerable DR (0.384) in the lineage leading to *Enterocytozoon bieneusi*, the only considerable large-scale duplication event found in the Opisthosporidia (9 species). *Gonapodya prolifera* presents the only detected large duplication event in Chytridiomycota (4 species), with a DR of 0.353. The base of Neocallimastigomycota presents the highest identified peak in the whole dataset, with a DR of 2.26. The group has been described to present a high expansions of certain enzyme families, such as the Carbohydrate Active Enzymes (CAZys), but the values obtained in this study cannot be explained by metabolic expansions alone. Even more, one of the two species represented in the group, *Pecoramyces ruminatus* presents a lineage-specific duplication peak (DR=0.572). However, a deeper look into the evolutionary history of this group is currently not possible due to the limited number of genomes from zoosporic fungi that are available. The two species represented in the Blastocladiomycota also present a duplication peak each, with a DR value of 1.057 for *Allomyces macroginus*, a well established polyploid; and 0.411 for *Catenaria anguillulae*. The Zoopagomycota also presents duplication peaks for *Basidiobolus meristosporus* (DR=0.828) and *Conidiobolus coronatus* (DR=0.314) in the Entomophthoromycotina (3 species); the base of Kickxellales (DR=0.337) and *Ramicandelaber brevisporus* (DR=0.403) for the Kickxellomycotina (4 species). Another peak appears in the only representative of Glomeromycota, *Rhizophagus irregularis* (DR=0.462). Mucoromycota, is the best represented non-Dikarya fungal clade, with 13 species. The group presents at least two well described WGD events (Corrochano *et al.* , 2016; Ma *et al.* , 2009) and both of them are recovered in our analysis. One of the events appears at the base of the order Mucorales (DR=0.484), while the other includes *Rhizopus delemar* (DR=0.271). On top of that, we find a peak in the branch subtending the two species of *Mortierella* (DR=0.727) included in our dataset; another at the very base of Mucoromycotina (DR=0.389); one more including both strains of *Rhizopus microsporus* (DR=0.505) and a last one in *Rhizopus microsporus* var. *chinensis* CCTCC M201021 (DR=0.304).

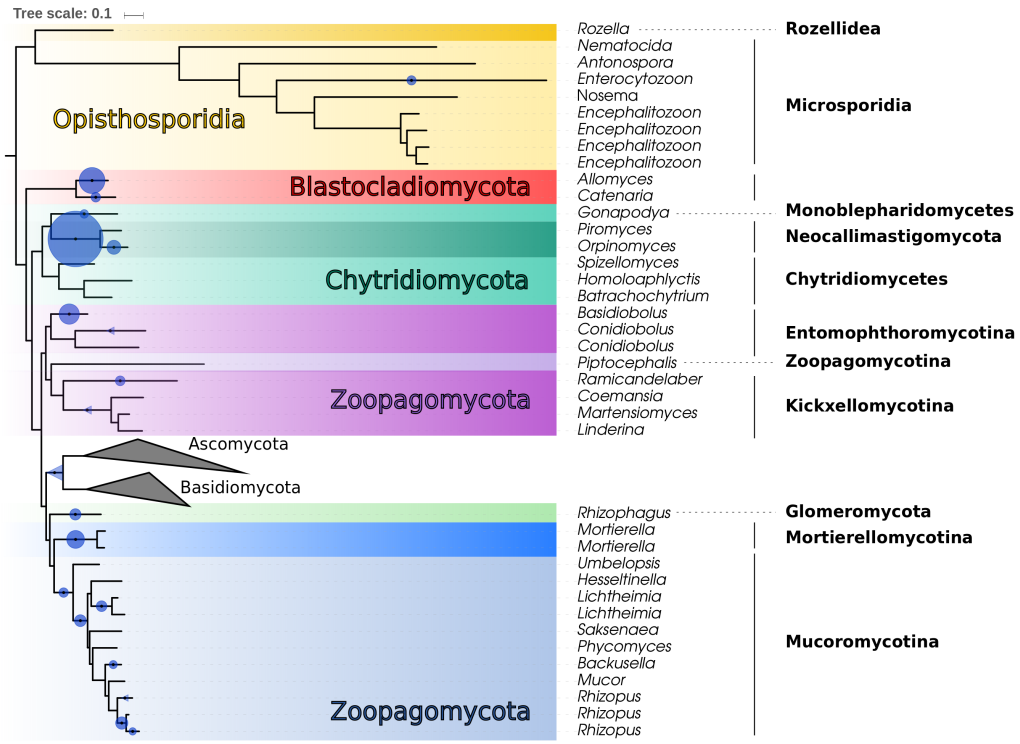

Figure 4. Duplication rates across non-Dikarya Fungi

Phylogenetic tree covering all represented Basidiomycota in the collection of phylomes. For the sake of simplicity, only the genus of each leaf is represented. Nodes with a duplication rate (DR) above 0.25 according to explorer calculations are marked with a symbol whose diameter is proportional to the value. Nodes with an corrected duplication rate (DR2) below 0.25 are marked with a triangle, otherwise they are marked with a circle. Tree visualization has been performed using the interactive Tree of Life server (iTOL) (Letunic & Bork, 2016)(Letunic & Bork, 2016).

Basidiomycota The duplication frequencies in this clade are shown in detail in figure 5. Duplication peaks seem to be absent in the Ustilaginomycotina. The group is known from having well conserved synteny and reduced genomes, due to their mostly plant biotrophic parasite or yeast-like lifestyles. The group is represented in our dataset by just eleven species, and thus we cannot discard that duplication peaks might arise in a larger dataset. In contrast, the situation in Pucciniomycotina, here represented by sixteen species is rather different. We find lineage-specific peaks in *Classicula fluitans* (DR=0.325) and *Atractiella rhizophila* (DR=0.651), each being the only representative of its respective classes. However, the main result is the extremely dynamic nature of the gene pool of Pucciniomycetes. This group, which presents three species in our dataset, has duplication peaks in each of the four included nodes. The result is not entirely unexpected, as these organisms, in contrast to the rest of the group, present extremely large genomes and gene numbers, and have very complex life cycles. The largest DR is present at the base of the genus *Puccinia* (DR=1.153). However, once corrected for gene expansions, the DR2 is just 0.367; the largest discrepancy in all our dataset. This indicates that rampant gene expansion, rather than WGD, is the most likely explanation for the detected duplication peaks and the high dynamism of genomes in the Pucciniomycetes (Tavares *et al.* , 2014).

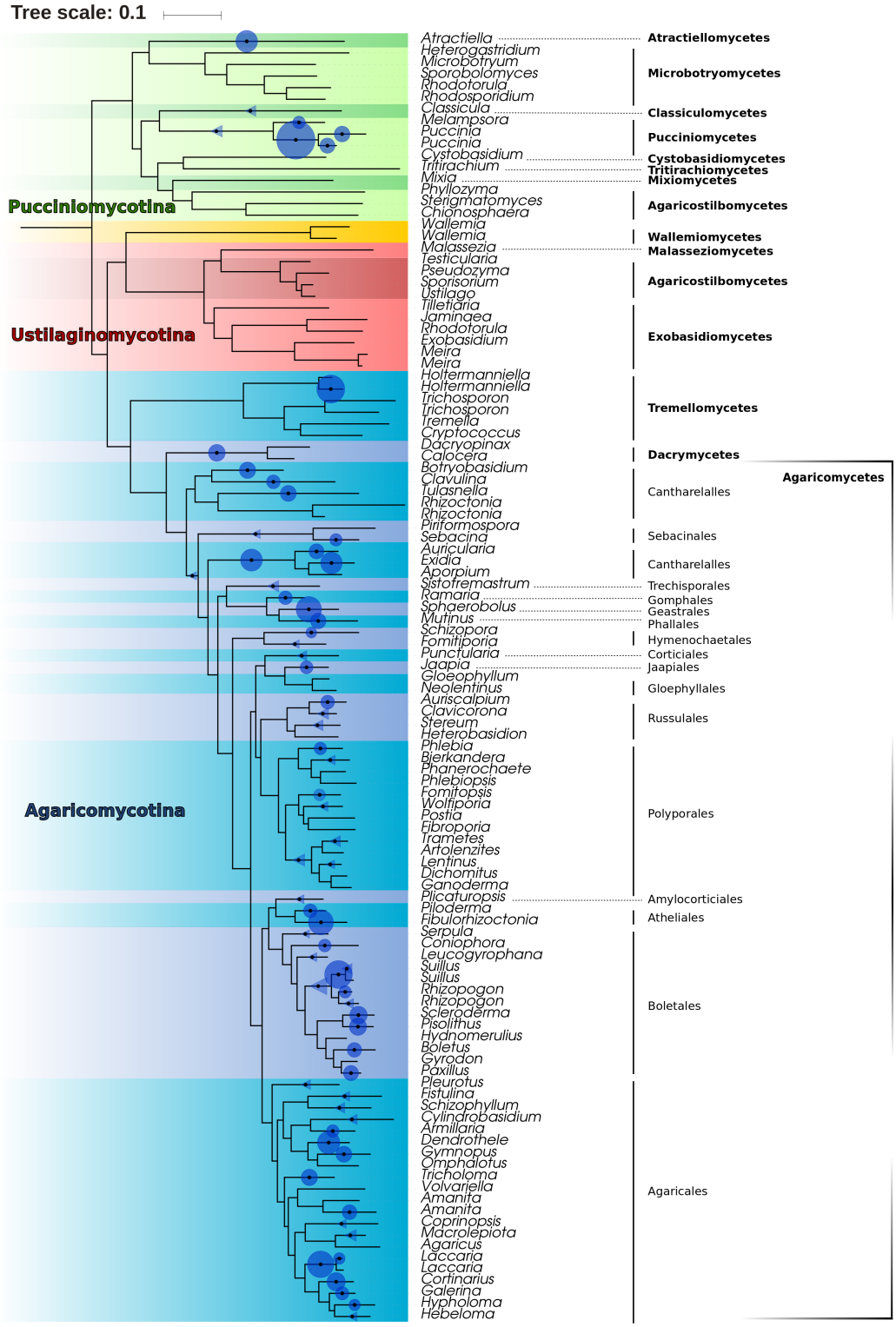

Figure 5. Duplication rates across Basidiomycota

Phylogenetic tree covering all represented non-Dikarya in the collection of phylomes. For the sake of simplicity, only the genus of each leaf is represented. Nodes with a duplication rate (DR) above 0.25 according to explorer calculations are marked with a symbol whose diameter is proportional to the value. Nodes with an corrected duplication rate (DR2) below 0.25 are marked with a triangle, otherwise they are marked with a circle. Tree visualization has been performed using the interactive Tree of Life server (iTOL) (Letunic & Bork, 2016).

The last group of Basidiomycota that we will discuss is the Agaricomycotina. For the sake of simplicity we will include the Wallemiomycetes, for which no duplication peaks are detected, as members of the Agaricomycotina, in concordance with our own phylogenetic results. In total, our dataset includes 84 species in the Agaricomycotina. From these, 51 species (60.71%) present duplication peaks with a DR above 0.25 at terminal nodes. Beyond these, there are several notable internal nodes, some of which we are highlighting. The base of Dacrymycetes presents a DR of 0.505, although this is a very ancient lineage represented by just 2 species. The Auriculariales (3 species) seem to have suffered a convoluted evolutionary history, with all individual species having lineage-specific duplications peaks and with an additional peak at the base of the order (DR=0.657). The Polyporaceae (5 species) is another clade that presents an ancestral duplication peak at its base (DR=0.370). The same applies to the Suillinae (4 species), a clade within Boletales represented here by the genera *Suillus* and *Rhizopogon* (DR=0.520), plus an additional peak at the base of the two species of *Suillus* (DR=0.832). Finally, there is another peak in the clade including the two species of the genus *Lacccaria* (DR=0.794).

More interesting, though, is an internal node that comprises all Agaricomycetes with a DR of 0.255. However, there is discrepancy with the results of phylome analyses. Phylomes 753 (*Agaricus bisporus*), 754 and 761 (*Fibroporia radiculosa*), 762 (*Auricularia subglabra*), 763 (*Ramaria acris*), 764 (*Hypholoma sublateritium*), 765 (*Paxillus involutus*) and 766 (*Ganoderma sp.*) identify a notable peak (DR of 0.322, 0.189, 0.356, 0.237, 0.497, 0.532 and 0.475; respectively) at the base of Agaricomycetes excluding Cantharellales and Sebacinales. This node is notable because it marks a clear delimitation on metabolic capabilities and fruiting body structure. The members of the clade affected by this ancient duplication peak comprise all homobasidiomycetes, that is, true complex fruiting-body forming fungi. The exception to these rules are the Auriculariales, which usually form simpler fruiting bodies; as well as *Clavulina* and *Botryobasidium* (Cantharellales), which produce complex fruiting bodies but are not included in the delimited clade. The clade also corresponds with the acquisition peak of lignin degrading enzymes at the end of the Carboniferous which has been previously described (Floudas *et al.* , 2012).

The phylome of *Auricularia subglabra* (phylome 762) reveals an extremely high duplication rate in Auriculariales. The branch including both *Auricularia* species (*A. subglabra* and *A. delicata*) has a duplication rate of 1.440. The node including both *Auricularia* species, *Exidia glandulosa* and *Elmeria caryae* has a duplication rate of 1.282. Both values are barely affected by the removal of expansions (1.435 and 1.279 respectively). However, the specific duplication rate of *Auricularia subglabra* is very low (0.062). Regarding the GO terms enriched in both highly duplicated nodes we find some commonalities: oxidoreductase activity, heme and iron binding, RNA-directed DNA polymerase activity, serine endopeptidase activity and structural components of the ribosome. While the enrichment on oxidative metabolism might reflect acquisition of lignin degrading capabilities in Agaricomycotina prior to the divergence of Auriculariales and perhaps independent evolution of the trait; the rest of the enriched term cannot easily be assigned to any phenotypic particularity of the order.

Ascomycota Duplication peaks in Ascomycota fungi are rare. Under the defined parameters, we only find seven of such peaks among 193 species. Within the Taphrinomycotina (12 species) there is only one peak, at the node including the two described members of the genus *Archaeorhizomyces*. Again, this is a very ancient lineage represented by only a handful of species. Saccharomycotina is an extremely well studied clade that includes 28 species in our dataset. Within it we find duplication peaks for *Millerozyma farinosa* (syn. *Pichia sorbitophila*) (DR=0.644), *Metschnikowia fructicola* (DR=0.344); and *Pichia kudriavzevii* (syn. *Candida krusei*) (DR=0.269). Both *M. farinosa* and *Met. fructicola* have been described as hybrid species, which could explain the high DR values. Such claim has not been made for *P. kudriavzevii*, though. Consistent with previous results, the node corresponding with the well described hybridization event in Saccharomycetaceae presents a DR of just 0.119 (Marcet-houben & Gabaldón, 2015). As mentioned above, duplication peaks are extremely rare within Pezizomycotina, with only three peaks detected in a dataset containing 153 species. The first and most obvious presents a DR of 0.925 and includes *Hortaea werneckii*, an extremely halotolerant black fungus that presents a well described WGD event involved in the expansion of its osmoprotective genetic repertoire (Lenassi *et al.* , 2013; Sinha *et al.* , 2017). The

second is a very small peak (DR=0.251) at the base of Orbiliomycetes, which includes two species in our dataset. The last peak, with a value of 0.785, correlates with the base of Leotiomyceta, the group including all Pezizomycotina excluding Orbiliomycetes and Pezizomycetes. However, DR values are not completely consistent across different phylomes, which we attribute to the different species sets and how deep the node is. This node deserves further exploration as it might be responsible for the largest diversification in the whole kingdom.

Discussion

Duplication peaks are a problematic way to look for WGD events for several reasons. Gene expansions might mislead estimations of DR. Accounting for those cases, we have developed corrective metrics that take into account this possibility. Even then, gene expansions are an important evolutionary phenomenon and it is still an important part of our results. DR estimates are based on phylome and thus for any given node it depends on its representativity in our phylome connection and the species distribution for each phylome. As such, while we have tried to make our best effort to cover the totality of the fungal tree of life, estimations for certain nodes in the tree might be based on a comparatively smaller dataset. Lastly, ancient lineages are difficult to interpret, as the duplications might have occurred either as an steady accumulation over time or through dramatic events such as WGD. Analogously, the event might have occurred in very recent times or hundreds of millions of years ago. This is particularly relevant for the non-Dikarya fungi, that are ancient lineages with a few species each. Even if such is the case, the results are consistent with known evidence in Mucorales, the only of these groups for which genomic representation could be argued to be adequate. Extrapolating these results towards an hypothetical adequate representation should give us a scenario not very different from what we have found in Agaricomycetes. The most surprising result is the extremely low frequency of duplication peaks in the Ascomycota, particularly in the Pezizomycotina. Curiously, this group presents the highest frequency of horizontal gene transfer of all Fungi. This phenomenon plays a much more central role in the acquisition of genetic novelties, too, perhaps indirectly reducing the necessity of gene duplication. Typically yeast-like clades seem to present few duplication peaks as well, with just a few examples in non yeast Pucciniomycotina and in well known hybrid lineages in Saccharomycotina. This last group presents a WGD event whose DR falls far below our thresholds. This might suggest that yeast-like lineages present strong evolutionary pressures favoring genome compaction that might reduce the signal of hypothetical events. If this is the case of Pezizomycotina, why is this not happening in Mucorales, that present very similar genome sizes? Another genomic peculiarity of Pezizomycotina is the existence of mesosynteny, which is the presence of genes that tend to appear in physical proximity across long evolutionary distances, but whose gene order is not conserved. Whether this might be correlated with our observed results needs to be investigated.

Table 1: List of phylomes and represented species

Seed species	Included species	PylomeID
Catenaria anguillulae PL171	*Puccinia graminis, Ustilago maydis, Coprinopsis cinerea, Yarrowia lypolytica, Arthrobotrys oligospora, Schizosaccharomyces pombe, Linderina pennispora, Coemansia reversa, Ramicandelaber brevisporus, Conidiobolus thromboides, Conidiobolus coronatus, Basidiobolus meristosporus, Batrachochytrium dendrobatidis, Spizellomyces punctatus, Gonapodya prolifera, Piromyces sp., Rhizophagus irregularis, Mortierella verticillata, Umbelopsis ramanniana, Catenaria anguillulae, Allomyces macrogynus, Rozella allomycis, Amphimedon queenslandica, Nematostella vectensis, Capsaspora owczarzaki, Monosiga brevicollis, Pythium ultimum*	416
Batrachochytrium dendrobatidis JAM81	*Puccinia graminis, Ustilago maydis, Coprinopsis cinerea, Yarrowia lypolytica, Arthrobotrys oligospora, Schizosaccharomyces pombe, Linderina pennispora, Coemansia reversa, Ramicandelaber brevisporus, Conidiobolus thromboides, Conidiobolus coronatus, Basidiobolus meristosporus, Batrachochytrium dendrobatidis, Spizellomyces punctatus, Gonapodya prolifera, Piromyces sp., Rhizophagus irregularis, Mortierella verticillata, Umbelopsis ramanniana, Catenaria anguillulae, Allomyces macrogynus, Rozella allomycis, Amphimedon queenslandica, Nematostella vectensis, Capsaspora owczarzaki, Monosiga brevicollis, Pythium ultimum*	417
Batrachochytrium dendrobatidis JAM81	*Orpinomyces sp, Gonapodya prolifera, Homoloaphlyctis polyrhiza, Amphimedon queenslandica, Puccinia graminis, Monosiga brevicollis, Nematostella vectensis, Allomyces macrogynus, Capsaspora owczarzaki, Spizellomyces punctatus, Piromyces sp. , Rhizophagus irregularis, Arthrobotrys oligospora, Catenaria anguillulae, Conidiobolus coronatus, Rozella allomycis*	419
Coemansia reversa	*Mortierella verticillata, Orpinomyces sp., Basidiobolus meristosporus, Conidiobolus thromboides, Gonapodya prolifera, Homoloaphlyctis polyrrhiza, Coprinopsis cinerea, Schizosaccharomyces pombe, Ramicandelaber brevisporus, Amphimedon queenslandica, Umbelopsis ramanniana, Puccinia graminis, Pythium ultimum, Monosiga brevicollis, Nematostella vectensis, Yarrowia lypolytica, Ustilago maydis, Allomyces macrogynus, Capsaspora owczarzaki, Linderina pennispora, Spizellomyces punctatus, Batrachochytrium dendrobatidis, Piromyces sp, Rhizophagus irregularis, Arthrobotrys oligospora, Coemansia reversa, Catenaria anguillulae, Conidiobolus coronatus, Rozella allomycis, Piptocephalis cylindrospora*	420
Choiromyces venosus 120613-1	*Pyronema omphalotes, Ascobolus immersus, Dactylellina haptotyla, Wilcoxina mikolae, Choiromyces venosus, Morchella conica, Aspergillus fumigatus, Botrytis cinerea, Ascodesmis nigricans, Mucor circinelloides, Mycosphaerella fijiensis, Saccharomyces cerevisiae, Sarcoscypha coccinea, Neurospora crassa, Agaricus bisporus, Saitoella complicata, Tuber melanosporum, Arthrobotrys oligospora, Terfezia boudieri*	665
Cadophora sp. DSE1049	*Ascocoryne sarcoides, Marssonina brunnea, Meliniomyces bicolor, Glarea lozoyensis, Meliniomyces varaibilis, Loramyces juncicola, Chalara longipes, Cadophora sp., Pseudographis elatina, Aspergillus fumigatus, Spathularia flavida, Botrytis cinerea, Mucor circinelloides, Mycosphaerella fijiensis, Thelebolus stercoreus, Saccharomyces cerevisiae, Neurospora crassa, Agaricus bisporus, Blumeria graminis, Saitoella complicata, Tuber melanosporum, Pseudogymnoascus destructans, Sclerotinia sclerotiorum, Hymenoscyphus fraxineus, Amorphotheca resinae, Oidiodendron maius*	666

Table 1: List of phylomes and represented species (Cont.)

Seed species	Included species	PylomeID
Thozetella sp. PMI_491	*Tolypocladium inflatum, Claviceps purpurea, Arthrinium arundis, Melanconium sp., Passalora fulva, Verticilium dahliae, Ophiostoma picea, Fusarium fujikuroi, Phaeoacremonium minimum, Eutypa lata, Ilionectria europaea, Microascus trigonosporus, Chaetosphaeria innumera, Anthostoma avocetta, Thozetella sp., Coniochaeta ligniaria, Magnaporthe oryzae, Niesslia exilis, Daldinia eschscholtzii, Clonostachys rosea, Chaetomium globosum, Calosphaeria pulchella, Aspergillus fumigatus, Botrytis cinerea, Epichloe festucae, Mucor circinelloides, Ophiostoma piliferum, Mycosphaerella fijiensis, Acremonium alcalophilum, Trichoderma virens, Trichoderma atroviride, Corollospora maritima, Colletotrichum gloeosporioides, Saccharomyces cerevisiae, Neurospora crassa, Podospora anserina, Agaricus bisporus, Gibberella zeae, Saitoella complicata, Thermothelomyces thermophila, Thielavia terrestris, Glomerella graminicola, Beauveria bassiana, Metarhizium anisopliae, Grosmannia clavigera, Tuber melanosporum, Nectria haematococca, Cryphonectria parasitica, Colletotrichum higginsianum, Sordaria macrospora, Cordyceps militaris*	667
Ilyonectria europaea PMI-82	*Tolypocladium inflatum, Claviceps purpurea, Arthrinium arundis, Melanconium sp., Passalora fulva, Verticillium dahliae, Ophiostoma piceae, Fusarium fujikuroi, Phaeoacremonium minimum, Eutypa lata, Ilyonectria europaea, Microascus trigonosporus, Chaetosphaeria innumera, Anthostoma avocetta, Thozetella sp., Coniochaeta ligniaria, Magnaporthe oryzae, Niesslia exilis, Daldinia eschscholtzii, Clonostachys rosea, Chaetomium globosum, Calosphaeria pulchella, Aspergillus fumigatus, Botrytis cinerea, Epichloe festucae, Mucor circinelloides, Ophiostoma piliferum, Mycosphaerella fijiensis, Acremonium alcalophilum, Trichoderma virens, Trichoderma atroviride, Corollospora maritima, Colletotrichum gloeosporioides, Saccharomyces cerevisiae, Neurospora crassa, Podospora anserina, Agaricus bisporus, Gibberella zeae, Saitoella complicata, Thermothelomyces thermophila, Thielavia terrestris, Glomerella graminicola, Beauveria bassiana, Metarhizium anisopliae, Grosmannia clavigera, Tuber melanosporum, Nectria haematococca, Cryphonectria parasitica, Colletotrichum higginsianum, Sordaria macrospora, Cordyceps militaris*	668
Graphis scripta	*Xanthoria parietina, Thermoascus aurantiacus, Gymnascella citrina, Monascus purpureus, Monascus ruber, Thermomyces lanuginosus, Capronia coronata, Trinosporium guianense, Cyphellophora europaea, Acarospora strigata, Xylona heveae, Cladonia grayi, Aspergillus fumigatus, Botrytis cinerea, Uncinocarpus reesii, Mucor circinelloides, Caliciopsis orientalis, Aspergillus niger, Mycosphaerella fijiensis, Ascosphaera apis, Endocarpon pallidulum, Talaromyces stipitatus, Penicillium marneffei, Saccharomyces cerevisiae, Penicillium rubens, Paracoccidioides brasiliensis, Ajellomyces capsulata, Neurospora crassa, Agaricus bisporus, Coccidioides immitis, Trichophyton rubrum, Saitoella complicata, Arthroderma benhamiae, Tuber melanosporum, Graphis scripta, Gymnascella aurantiaca, Dibaeis baeomyces, Exophiala dermatitidis, Symbiotaphrina kochii*	669
Gymnascella citrina NRRL 5970	*Xanthoria parietina, Thermoascus aurantiacus, Gymnascella citrina, Monascus purpureus, Monascus ruber, Thermomyces lanuginosus, Capronia coronata, Trinosporium guianense, Cyphellophora europaea, Acarospora strigata, Xylona heveae, Cladonia grayi, Aspergillus fumigatus, Botrytis cinerea, Uncinocarpus reesii, Mucor circinelloides, Caliciopsis orientalis, Aspergillus niger, Mycosphaerella fijiensis, Ascosphaera apis, Endocarpon pallidulum, Talaromyces stipitatus, Penicillium marneffei, Saccharomyces cerevisiae, Penicillium rubens, Paracoccidioides brasiliensis, Ajellomyces capsulata, Neurospora crassa, Agaricus bisporus, Coccidioides immitis, Trichophyton rubrum, Saitoella complicata, Arthroderma benhamiae, Tuber melanosporum, Graphis scripta, Gymnascella aurantiaca, Dibaeis baeomyces, Exophiala dermatitidis, Symbiotaphrina kochii*	670

Table 1: List of phylomes and represented species (Cont.2)

Seed species	Included species	PylomeID
Pyrenophora tritici-repentis	*Aureobasidium pullulans, Mycosphaerella graminicola, Tothia fuscella, Zasmidium cellare, Botryosphaeria dothidea, Hysterium pulicare, Didymella exigua, Cladosporium sphaerospermum, Passalora fulva, Cucurbitaria berberidis, Lentithecium fluviatile, Myriangium duriaei, Aplosporella prunicola, Aulographum hederae, Polychaeton citri, Lepidopterella palustris, Zopfia rhizophila, Piedraia hortae, Dissoconium aciculare, Saccharata proteae, Lophiostoma macrostomum, Pleomassaria siparia, Melanomma pulvis-pyrius, Patellaria atrata, Eremomyces bilateralis, Dothidotthia symphoricarpi, Amniculicola lignicola, Karstenula rhodostoma, Lophium mytilinum, Bimuria novae-zelandiae, Trematosphaeria pertusa, Byssothecium circinans, Delitschia confertaspora, Phaeosphaeria nodorum, Aspergillus fumigatus, Botrytis cinerea, Mucor circinelloides, Trypethelium eluteriae, Rhytidhysteron rufulum, Mycosphaerella fijiensis, Pyrenophora tritici-repentis, Baudoinia compniacensis, Saccharomyces cerevisiae, Cochliobolus heterostrophus, Neurospora crassa, Agaricus bisporus, Saitoella complicata, Glonium stellatum, Corynespora cassiicola, Tuber melanosporum, Dothistroma septosporum, Polyplosphaeria fusca, Sphaerulina musiva, Trichodelitschia bisporula, Clathrospora elynae, Sporormia fimetaria, Acidomyces richmondensis, Cenococcum geophilum, Hortaea werneckii, Periconia macrospinosa, Zymoseptoria pseudotritici, Leptosphaeria maculans*	671
Passalora fulva CBS 131901	*Aureobasidium pullulans, Mycosphaerella graminicola, Tothia fuscella, Zasmidium cellare, Botryosphaeria dothidea, Hysterium pulicare, Didymella exigua, Cladosporium sphaerospermum, Passalora fulva, Cucurbitaria berberidis, Lentithecium fluviatile, Myriangium duriaei, Aplosporella prunicola, Aulographum hederae, Polychaeton citri, Lepidopterella palustris, Zopfia rhizophila, Piedraia hortae, Dissoconium aciculare, Saccharata proteae, Lophiostoma macrostomum, Pleomassaria siparia, Melanomma pulvis-pyrius, Patellaria atrata, Eremomyces bilateralis, Dothidotthia symphoricarpi, Amniculicola lignicola, Karstenula rhodostoma, Lophium mytilinum, Bimuria novae-zelandiae, Trematosphaeria pertusa, Byssothecium circinans, Delitschia confertaspora, Phaeosphaeria nodorum, Aspergillus fumigatus, Botrytis cinerea, Mucor circinelloides, Trypethelium eluteriae, Rhytidhysteron rufulum, Mycosphaerella fijiensis, Pyrenophora tritici-repentis, Baudoinia compniacensis, Saccharomyces cerevisiae, Cochliobolus heterostrophus, Neurospora crassa, Agaricus bisporus, Saitoella complicata, Glonium stellatum, Corynespora cassiicola, Tuber melanosporum, Dothistroma septosporum, Polyplosphaeria fusca, Sphaerulina musiva, Trichodelitschia bisporula, Clathrospora elynae, Sporormia fimetaria, Acidomyces richmondensis, Cenococcum geophilum, Hortaea werneckii, Periconia macrospinosa, Zymoseptoria pseudotritici, Leptosphaeria maculans*	672
Aspergillus fumigatus	*Aureobasidium pullulans, Tothia fuscella, Archaeorhizomyces finlayi, Pyronema omphalodes, Dactylellina haptotyla, Eutypa lata, Xylona heveae, Archaeorhizomyces borealis, Magnaporthe oryzae, Kluyveromyces lactis, Kuraishia capsulata, Aspergillus fumigatus, Mucor circinelloides, Botrytis fuckeliana, Blastobotrys adeninivorans, Saccharomyces cerevisiae, Yarrowia lipolytica, Neurospora crassa, Agaricus bisporus, Candida albicans, Saitoella complicata, Tuber melanosporum, Pseudogymnoascus destructans, Melampsora laricis-populina, Arthrobotrys oligospora*	750
Neurospora crassa	*Aureobasidium pullulans, Tothia fuscella, Archaeorhizomyces finlayi, Pyronema omphalodes, Dactylellina haptotyla, Eutypa lata, Xylona heveae, Archaeorhizomyces borealis, Magnaporthe oryzae, Kluyveromyces lactis, Kuraishia capsulata, Aspergillus fumigatus, Mucor circinelloides, Botrytis fuckeliana, Blastobotrys adeninivorans, Saccharomyces cerevisiae, Yarrowia lipolytica, Neurospora crassa, Agaricus bisporus, Candida albicans, Saitoella complicata, Tuber melanosporum, Pseudogymnoascus destructans, Melampsora laricis-populina, Arthrobotrys oligospora*	751

Table 1: List of phylomes and represented species (Cont.3)

Seed species	Included species	PylomeID
Arthrobotrys oligospora ATCC 24927	*Aureobasidium pullulans, Tothia fuscella, Archaeorhizomyces finlayi, Pyronema omphalodes, Dactylellina haptotyla, Eutypa lata, Xylona heveae, Archaeorhizomyces borealis, Magnaporthe oryzae, Kluyveromyces lactis, Kuraishia capsulata, Aspergillus fumigatus, Mucor circinelloides, Botrytis fuckeliana, Blastobotrys adeninivorans, Saccharomyces cerevisiae, Yarrowia lipolytica, Neurospora crassa, Agaricus bisporus, Candida albicans, Saitoella complicata, Tuber melanosporum, Pseudogymnoascus destructans, Melampsora laricis-populina, Arthrobotrys oligospora*	752
Agaricus bisporus	*Fibroporia radiculosa, Piriformospora indica, Pleurotus ostreatus, Wallemia ichthyophaga, Exobasidium vaccinii, Heterogastridium pycnidioideum, Cryptococcus neoformans, Aspergillus fumigatus, Mucor circinelloides, Puccinia graminis, Malassezia globosa, Neurospora crassa, Tremella mesenterica, Ustilago maydis, Agaricus bisporus, Saitoella complicata, Serpula lacrymans, Gloeophylum trabeum, Wallemia sebi, Melampsora laricis-populina, Mixia osmundae, Rhizoctonia solani*	753
Ustilago maydis	*Tilletiaria anomala, Rhodotorula phylloplana, Moesziomyces antarcticus, Meira miltonrushii, Exobasidium vaccinii, Jaminaea sp., Testicularia cyperi, Meira geulakonigii, Aspergillus fumigatus, Mucor circinelloides, Malassezia globosa, Saccharomyces cerevisiae, Cryptococcus neoformans, Ustilago maydis, Agaricus bisporus, Saitoella complicata, Batrachochytrium dendrobatidis, Melampsora laricis-populina, Mixia osmundae, Sporisorium reilianum*	755
Exobasidium vaccinii	*Tilletiaria anomala, Rhodotorula phylloplana, Moesziomyces antarcticus, Meira miltonrushii, Exobasidium vaccinii, Jaminaea sp., Testicularia cyperi, Meira geulakonigii, Aspergillus fumigatus, Mucor circinelloides, Malassezia globosa, Saccharomyces cerevisiae, Cryptococcus neoformans, Ustilago maydis, Agaricus bisporus, Saitoella complicata, Batrachochytrium dendrobatidis, Melampsora laricis-populina, Mixia osmundae, Sporisorium reilianum*	756
Rhodotorula graminis	*Tritirachium sp., Atractiellales sp., Exobasidium vaccinii, Phyllozyma linderae, Heterogastridium pycnidioideum, Classicula fluitans, Rhodotorula graminis, Rhodotorula minuta, Aspergillus fumigatus, Mucor circinelloides, Sporobolomyces roseus, Puccinia graminis, Sterigmatomyces hyphaenes, Saccharomyces cerevisiae, Cryptococcus neoformans, Ustilago maydis, Microbotryum violaceum, Rhodosporidium toruloides, Agaricus bisporus, Chionosphaera apobasidialis, Saitoella complicata, Batrachochytrium dendrobatidis, Melampsora laricis-populina, Mixia osmundae, Puccinia striiformis*	757
Melampsora laricis-populina 98AG31	*Tritirachium sp., Atractiellales sp., Exobasidium vaccinii, Phyllozyma linderae, Heterogastridium pycnidioideum, Classicula fluitans, Rhodotorula graminis, Rhodotorula minuta, Aspergillus fumigatus, Mucor circinelloides, Sporobolomyces roseus, Puccinia graminis, Sterigmatomyces hyphaenes, Saccharomyces cerevisiae, Cryptococcus neoformans, Ustilago maydis, Microbotryum violaceum, Rhodosporidium toruloides, Agaricus bisporus, Chionosphaera apobasidialis, Saitoella complicata, Batrachochytrium dendrobatidis, Melampsora laricis-populina, Mixia osmundae, Puccinia striiformis*	758

Table 1: List of phylomes and represented species (Cont.4)

Seed species	Included species	PylomeID
Mixia osmundae IAM 14324	*Tritirachium sp. Atractiellales sp., Exobasidium vaccinii, Phyllozyma linderae, Heterogastridium pycnidioideum, Classicula fluitans, Rhodotorula graminis, Rhodotorula minuta, Aspergillus fumigatus, Mucor circinelloides, Sporobolomyces roseus, Puccinia graminis, Sterigmatomyces hyphaenes, Saccharomyces cerevisiae, Cryptococcus neoformans, Ustilago maydis, Microbotryum violaceum, Rhodosporidium toruloides, Agaricus bisporus, Chionosphaera apobasidialis, Saitoella complicata, Batrachochytrium dendrobatidis, Melampsora laricis-populina, Mixia osmundae, Puccinia striiformis*	759
Wallemia sebi CBS 633.66	*Mortierella verticillata, Fibroporia radiculosa, Wallemia ichthyophaga, Rhodotorula graminis, Mucor circinelloides, Malassezia globosa, Saccharomyces cerevisiae, Neurospora crassa, Cryptococcus neoformans, Ustilago maydis, Agaricus bisporus, Gloeophylum trabeum, Wallemia sebi, Rhizophagus irregularis, Melampsora laricis-populina, Mixia osmundae, Rhizoctonia solani*	760
Auricularia subglabra TFB-10046 SS5	*Tulasnella calospora, Fibroporia radiculosa, Elmerina caryae, Trichosporon asahii, Ramaria acris, Sistotremastrum suecicum, Exidia glandulosa, Auricularia delicata, Cryptococcus neoformans, Botryobasidium botryosum, Schizopora paradoxa, Mutinus elegans, Tremella mesenterica, Ustilago maydis, Agaricus bisporus, Dacryopinax sphenocarpa, Gloeophylum trabeum, Fomitiporia mediterranea, Auricularia subglabra, Mixia osmundae, Serendipita vermifera, Rhizoctonia solani, Sphaerobolus stellatus*	762
Ramaria acris UT-36052-T	*Tulasnella calospora, Piriformospora indica, Elmerina caryae, Wallemia ichthyophaga, Ramaria acris, Sistotremastrum suecicum, Exidia glandulosa, Boletus edulis, Calocera cornea, Clavulina sp., Auricularia delicata, Coprinopsis cinerea, Botryobasidium botryosum, Schizopora paradoxa, Mutinus elegans, Aspergillus fumigatus, Cryptococcus neoformans, Tremella mesenterica, Ustilago maydis, Dacryopinax sphenocarpa, Gloeophylum trabeum, Wallemia sebi, Fomitiporia mediterranea, Auricularia subglabra, Mixia osmundae, Trichosporon asahii, Cutaneotrichosporon oleaginosus, Serendipita vermifera, Rhizoctonia solani, Holtermanniella takashimae, Sphaerobolus stellatus, Holtermanniella mycelialis*	763
Hypholoma sublateritium	*Fibroporia radiculosa, Tricholoma matsutake, Laccaria amethystina, Galerina marginata, Omphalotus olearius, Fistulina hepatica, Ramaria rubella, Pleurotus ostreatus, Cortinarius glaucopus, Cylindrobasidium torrendii, Dendrothele bispora, Armillaria mellea, Boletus edulis, Macrolepiota fuliginosa, Gymnopus luxurians, Coprinopsis cinerea, Schizopora paradoxa, Laccaria bicolor, Tremella mesenterica, Schizophylum commune, Agaricus bisporus, Serpula lacrymans, Dacryopinax sphenocarpa, Gloeophylum trabeum, Wallemia sebi, Amanita thiersii, Volvariella volvacea, Hypholoma sublateritium, Hebeloma cylindrosporum, Amanita muscaria, Rhizoctonia solani*	764
Paxillus involutus	*Scleroderma citrinum, Fibroporia radiculosa, Ramaria rubella, Pleurotus ostreatus, Rhizopogon vinicolor, Boletus edulis, Plicaturopsis crispa, Gyrodon lividus, Schizopora paradoxa, Laccaria bicolor, Hydnomerulius pinastri, Fibularhizoctonia sp., Suillus brevipes, Tremella mesenterica, Serpula lacrymans, Dacryopinax sphenocarpa, Rhizopogon salebrosus, Gloeophylum trabeum, Wallemia sebi, Paxillus involutus, Piloderma croceum, Coniophora puteana, Leucogyrophana mollusca, Pisolithus tinctorius, Suillus luteus, Rhizoctonia solani*	765

Table 1: List of phylomes and represented species (Cont.5)

Seed species	Included species	PylomeID
Ganoderma sp. 10597 SS1	*Fibroporia radiculosa, Ramaria rubella, Pleurotus ostreatus, Neolentinus lepideus, Boletus edulis, Lentinus tigrinus, Phlebia brevispora, Jaapia argillacea, Schizopora paradoxa, Laccaria bicolor, Auriscalpium vulgare, Artomyces pyxidatus, Tremella mesenterica, Trametes elegans, Phanerochaete chrysosporium, Trametes versicolor, Bjerkandera adusta, Postia placenta, Serpula lacrymans, Dacryopinax sphenocarpa, Gloeophylum trabeum, Wallemia sebi, Stereum hirsutum, Dichomitus squalens, Punctularia strigosozonata, Fomitopsis pinicola, Heterobasidion irregulare, Ganoderma sp., Wolfiporia cocos, Phlebiopsis gigantea, Rhizoctonia solani*	766
Schizosaccharomyces pombe (strain 972/ATCC 24843)	*Pneumocystis murina, Archaeorhizomyces finlayi, Taphrina deformans, Batrachochytrium dendrobatidis, Pneumocystis jirovecii, Archaeorhizomyces borealis, Schizosaccharomyces pombe, Aspergillus fumigatus, Mucor circinelloides, Pneumocystis carinii, Neolecta irregularis, Schizosaccharomyces japonicus, Schizosaccharomyces octosporus, Saccharomyces cerevisiae, Yarrowia lipolytica, Neurospora crassa, Ustilago maydis, Agaricus bisporus, Saitoella complicata, Schizosaccharomyces cryophilus*	767
Metschnikowia fructicola 277	*Tetrapisispora phaffii, Kazachstania africana, Torulaspora delbrueckii, Metschnikowia fructicola, Emericella nidulans, Ashbya gossypii, Schizosaccharomyces pombe, Kluyveromyces lactis, Kuraishia capsulata, Mucor circinelloides, Clavispora lusitaniae, Lodderomyces elongisporus, Blastobotrys adeninivorans, Cephaloascus albidus, Pichia kudriavzevii, Millerozyma farinosa, Pichia pastoris, Pichia stipitis, Pichia guilliermondii, Saccharomyces cerevisiae, Yarrowia lipolytica, Zygosaccharomyces rouxii, Debaryomyces hansenii, Dekkera bruxellensis, Neurospora crassa, Nakaseomyces delphensis, Nakaseomyces bacillisporus, Starmera amethionina, Candida albicans, Spathaspora passalidarum, Ogataea parapolymorpha, Cyberlindnera jadinii*	768
Cephaloascus albidus	*Tetrapisispora phaffii, Kazachstania africana, Torulaspora delbrueckii, Metschnikowia fructicola, Emericella nidulans, Ashbya gossypii, Schizosaccharomyces pombe, Kluyveromyces lactis, Kuraishia capsulata, Mucor circinelloides, Clavispora lusitaniae, Lodderomyces elongisporus, Blastobotrys adeninivorans, Cephaloascus albidus, Pichia kudriavzevii, Millerozyma farinosa, Pichia pastoris, Pichia stipitis, Pichia guilliermondii, Saccharomyces cerevisiae, Yarrowia lipolytica, Zygosaccharomyces rouxii, Debaryomyces hansenii, Dekkera bruxellensis, Neurospora crassa, Nakaseomyces delphensis, Nakaseomyces bacillisporus, Starmera amethionina, Candida albicans, Spathaspora passalidarum, Ogataea parapolymorpha, Cyberlindnera jadinii*	769
Cyberlindnera jadinii NRRL Y-1542	*Tetrapisispora phaffii, Kazachstania africana, Torulaspora delbrueckii, Metschnikowia fructicola, Emericella nidulans, Ashbya gossypii, Schizosaccharomyces pombe, Kluyveromyces lactis, Kuraishia capsulata, Mucor circinelloides, Clavispora lusitaniae, Lodderomyces elongisporus, Blastobotrys adeninivorans, Cephaloascus albidus, Pichia kudriavzevii, Millerozyma farinosa, Pichia pastoris, Pichia stipitis, Pichia guilliermondii, Saccharomyces cerevisiae, Yarrowia lipolytica, Zygosaccharomyces rouxii, Debaryomyces hansenii, Dekkera bruxellensis, Neurospora crassa, Nakaseomyces delphensis, Nakaseomyces bacillisporus, Starmera amethionina, Candida albicans, Spathaspora passalidarum, Ogataea parapolymorpha, Cyberlindnera jadinii*	770

References

Altschul, Stephen F and Gish, Warren and Miller, Webb, Eugene W. Myers, & Lipman, David J. 1990. *BLAST*.

Bauer, Robert, Garnica, Sigisfredo, Oberwinkler, Franz, Riess, Kai, Weiss, Michael, & Begerow, Dominik. 2015. Entorrhizomycota: A New Fungal Phylum Reveals New Perspectives on the Evolution of Fungi. *PloS one*, **10**(7).

Campbell, Matthew A., Ganley, Austen R. D., Gabaldón, Toni, & Cox, Murray P. 2016. The Case of the Missing Ancient Fungal Polyploids. *The American Naturalist*, **188**(6), 602–614.

Capella-Gutiérrez, Salvador, Silla-Martínez, José M, & Gabaldón, Toni. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**(15), 1972–3.

Corrochano, Luis M, Kuo, Alan, Marcet-Houben, Marina, Polaino, Silvia, Salamov, Asaf, Villalobos-Escobedo, José M, Grimwood, Jane, Álvarez, M Isabel, Avalos, Javier, Bauer, Diane, Benito, Ernesto P, Benoit, Isabelle, Burger, Gertraud, Camino, Lola P, Cánovas, David, Cerdá-Olmedo, Enrique, Cheng, Jan-Fang, Domínguez, Angel, Eliáš, Marek, Eslava, Arturo P, Glaser, Fabian, Gutiérrez, Gabriel, Heitman, Joseph, Henrissat, Bernard, Iturriaga, Enrique A, Lang, B Franz, Lavín, José L, Lee, Soo Chan, Li, Wenjun, Lindquist, Erika, López-García, Sergio, Luque, Eva M, Marcos, Ana T, Martin, Joel, McCluskey, Kevin, Medina, Humberto R, Miralles-Durán, Alejandro, Miyazaki, Atsushi, Muñoz-Torres, Elisa, Oguiza, José A, Ohm, Robin A, Olmedo, María, Orejas, Margarita, Ortiz-Castellanos, Lucila, Pisabarro, Antonio G, Rodríguez-Romero, Julio, Ruiz-Herrera, José, Ruiz-Vázquez, Rosa, Sanz, Catalina, Schackwitz, Wendy, Shahriari, Mahdi, Shelest, Ekaterina, Silva-Franco, Fátima, Soanes, Darren, Syed, Khajamohiddin, Tagua, Víctor G, Talbot, Nicholas J, Thon, Michael R, Tice, Hope, de Vries, Ronald P, Wiebenga, Ad, Yadav, Jagjit S, Braun, Edward L, Baker, Scott E, Garre, Victoriano, Schmutz, Jeremy, Horwitz, Benjamin A, Torres-Martínez, Santiago, Idnurm, Alexander, Herrera-Estrella, Alfredo, Gabaldón, Toni, & Grigoriev, Igor V. 2016. Expansion of Signal Transduction Pathways in Fungi by Extensive Genome Duplication. *Current biology*, **26**(12), 1577–1584.

Edgar, Robert C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, **32**(5), 1792–7.

Floudas, Dimitrios, Binder, Manfred, Riley, Robert, Barry, Kerrie, Blanchette, Robert A, Henrissat, Bernard, Martínez, Angel T, Otillar, Robert, Spatafora, Joseph W, Yadav, Jagjit S, Aerts, Andrea, Benoit, Isabelle, Boyd, Alex, Carlson, Alexis, Copeland, Alex, Coutinho, Pedro M, de Vries, Ronald P, Ferreira, Patricia, Findley, Keisha, Foster, Brian, Gaskell, Jill, Glotzer, Dylan, Górecki, Paweł, Heitman, Joseph, Hesse, Cedar, Hori, Chiaki, Igarashi, Kiyohiko, Jurgens, Joel A, Kallen, Nathan, Kersten, Phil, Kohler, Annegret, Kües, Ursula, Kumar, T K Arun, Kuo, Alan, LaButti, Kurt, Larrondo, Luis F, Lindquist, Erika, Ling, Albee, Lombard, Vincent, Lucas, Susan, Lundell, Taina, Martin, Rachael, McLaughlin, David J, Morgenstern, Ingo, Morin, Emanuelle, Murat, Claude, Nagy, Laszlo G, Nolan, Matt, Ohm, Robin A, Patyshakuliyeva, Aleksandrina, Rokas, Antonis, Ruiz-Dueñas, Francisco J, Sabat, Grzegorz, Salamov, Asaf, Samejima, Masahiro, Schmutz, Jeremy, Slot, Jason C, St John, Franz, Stenlid, Jan, Sun, Hui, Sun, Sheng, Syed, Khajamohiddin, Tsang, Adrian, Wiebenga, Ad, Young, Darcy, Pisabarro, Antonio, Eastwood, Daniel C, Martin, Francis, Cullen, Dan, Grigoriev, Igor V, & Hibbett, David S. 2012. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science (New York, N.Y.)*, **336**(6089), 1715–9.

Gabaldón, Toni. 2008. Large-scale assignment of orthology: back to phylogenetics? *Genome Biology*, **9**(10), 235.

Gascuel, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution*, **14**(0737-4038 (Print)), 685–695.

Guindon, Stéphane, Dufayard, Jean-François, Lefort, Vincent, Anisimova, Maria, Hordijk, Wim, & Gascuel, Olivier. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, **59**(3), 307–21.

Huerta-Cepas, J., & Gabaldon, T. 2011. Assigning duplication events to relative temporal scales in genome-wide studies. *Bioinformatics*, **27**(1), 38–45.

Huerta-Cepas, Jaime, & Gabaldón, Toni. 2011. Assigning duplication events to relative temporal scales in genome-wide studies. *Bioinformatics*, **27**(1), 38–45.

Huerta-Cepas, Jaime, Capella-Gutierrez, Salvador, Pryszcz, Leszek P, Denisov, Ivan, Kormes, Diego, Marcet-Houben, Marina, & Gabaldón, Toni. 2011. PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. *Nucleic acids research*, **39**(Database issue), D556–60.

Huerta-Cepas, Jaime, Serra, François, & Bork, Peer. 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Molecular Biology and Evolution*, **33**(6), 1635–1638.

James, Timothy Y., Porter, Teresita M., & Martin, W. Wallace. 2014. 7 Blastocladiomycota. *Pages 177–207 of: Systematics and Evolution*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Katoh, Kazutaka, & Standley, Daron M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, **30**(4), 772–80.

Katoh, Kazutaka, Kuma, Kei-ichi, Toh, Hiroyuki, & Miyata, Takashi. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, **33**(2), 511–518.

Kuramae, Eiko E., Robert, Vincent, Snel, Berend, & Boekhout, Teun. 2006. Conflicting phylogenetic position of Schizosaccharomyces pombe. *Genomics*, **88**(4), 387–393.

Landan, G., & Graur, D. 2007. Heads or Tails: A Simple Reliability Check for Multiple Sequence Alignments. *Molecular Biology and Evolution*, **24**(6), 1380–1383.

Lassmann, Timo, & Sonnhammer, Erik L L. 2005. Kalign–an accurate and fast multiple sequence alignment algorithm. *BMC bioinformatics*, **6**, 298.

Le, S. Q., & Gascuel, O. 2008. An Improved General Amino Acid Replacement Matrix. *Molecular Biology and Evolution*, **25**(7), 1307–1320.

Lenassi, Metka, Gostinčar, Cene, Jackman, Shaun, Turk, Martina, Sadowski, Ivan, Nislow, Corey, Jones, Steven, Birol, Inanc, Cimerman, Nina Gunde, & Plemenitaš, Ana. 2013. Whole Genome Duplication and Enrichment of Metal Cation Transporters Revealed by De Novo Genome Sequencing of Extremely Halotolerant Black Yeast Hortaea werneckii. *PLoS ONE*, **8**(8), e71328.

Letunic, Ivica, & Bork, Peer. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research*, **44**(W1), W242–W245.

Liu, Y., Leigh, J. W., Brinkmann, H., Cushion, M. T., Rodriguez-Ezpeleta, N., Philippe, H., & Lang, B. F. 2008. Phylogenomic Analyses Support the Monophyly of Taphrinomycotina, including Schizosaccharomyces Fission Yeasts. *Molecular Biology and Evolution*, **26**(1), 27–34.

Ma, Li-Jun, Ibrahim, Ashraf S, Skory, Christopher, Grabherr, Manfred G, Burger, Gertraud, Butler, Margi, Elias, Marek, Idnurm, Alexander, Lang, B Franz, Sone, Teruo, Abe, Ayumi, Calvo, Sarah E, Corrochano, Luis M, Engels, Reinhard, Fu, Jianmin, Hansberg, Wilhelm, Kim, Jung-Mi, Kodira, Chinnappa D, Koehrsen, Michael J, Liu, Bo, Miranda-Saavedra, Diego, O'Leary, Sinead, Ortiz-Castellanos, Lucila, Poulter, Russell, Rodriguez-Romero, Julio, Ruiz-Herrera, José, Shen, Yao-Qing, Zeng, Qiandong, Galagan, James, Birren, Bruce W, Cuomo, Christina A, & Wickes, Brian L. 2009. Genomic analysis of the basal lineage fungus Rhizopus oryzae reveals a whole-genome duplication. *PLoS genetics*, **5**(7), e1000549.

Marcet-houben, Marina, & Gabaldón, Toni. 2015. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. *PLoS biology*, **13**(8), 1–26.

Matheny, P. Brandon, Gossmann, Jasmin A., Zalar, Polona, Kumar, T.K. Arun, & Hibbett, David S. 2006. Resolving the phylogenetic position of the Wallemiomycetes: an enigmatic major lineage of Basidiomycota. *Canadian Journal of Botany*, **84**(12), 1794–1805.

Menkis, Audrius, Urbina, Hector, James, Timothy Y., & Rosling, Anna. 2014. Archaeorhizomyces borealis sp. nov. and a sequence-based classification of related soil fungal species. *Fungal Biology*, **118**(12), 943–955.

Mishra, Bagdevi, Choi, Young-Joon, & Thines, Marco. 2018. Phylogenomics of Bartheletia paradoxa reveals its basal position in Agaricomycotina and that the early evolutionary history of basidiomycetes was rapid and probably not strictly bifurcating. *Mycological Progress*, **17**(3), 333–341.

Nguyen, Hai D T, Chabot, Denise, Hirooka, Yuuri, Roberson, Robert W, & Seifert, Keith A. 2015a. Basidioascus undulatus : genome, origins and sexuality. *International Microbiological Association Fungus*, **6**(1), 215–231.

Padamsee, Mahajabeen, Arun Kumar, T K, Riley, Robert, Binder, Manfred, Boyd, Alex, Calvo, Ana M, Furukawa, Kentaro, Hesse, Cedar, Hohmann, Stefan, James, Tim Y, Labutti, Kurt, Lapidus, Alla, Lindquist, Erika, Lucas, Susan, Miller, Kari, Shantappa, Sourabha, Grigoriev, Igor V, Hibbett, David S, Mclaughlin, David J, Spatafora, Joseph W, & Aime, M Catherine. 2012. The genome of the xerotolerant mold Wallemia sebi reveals adaptations to osmotic stress and suggests cryptic sexual reproduction. *Fungal Genetics and Biology*, **49**, 217–226.

Price, Morgan N., Dehal, Paramvir S., & Arkin, Adam P. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE*, **5**(3).

Q.-M. Wang, Theelen, B., M. Groenewald, F.-Y. Bai, & T. Boekhout. 2014. Moniliellomycetes and Malasseziomycetes , two new classes in Ustilaginomycotina. *Persoonia*, **33**, 41–47.

Schoch, Conrad L., Sung, Gi-Ho, López-Giráldez, Francesc, Townsend, Jeffrey P., Miadlikowska, Jolanta, Hofstetter, Valérie, Robbertse, Barbara, Matheny, P. Brandon, Kauff, Frank, Wang, Zheng, Gueidan, Cécile, Andrie, Rachael M., Trippe, Kristin, Ciufetti, Linda M., Wynns, Anja, Fraker, Emily, Hodkinson, Brendan P., Bonito, Gregory, Groenewald, Johannes Z., Arzanlou, Mahdi, Sybren de Hoog, G., Crous, Pedro W., Hewitt, David, Pfister, Donald H., Peterson, Kristin, Gryzenhout, Marieka, Wingfield, Michael J., Aptroot, André, Suh, Sung-Oui, Blackwell, Meredith, Hillis, David M., Griffith, Gareth W., Castlebury, Lisa A., Rossman, Amy Y., Lumbsch, H. Thorsten, Lücking, Robert, Büdel, Burkhard, Rauhut, Alexandra, Diederich, Paul, Ertz, Damien, Geiser, David M., Hosaka, Kentaro, Inderbitzin, Patrik, Kohlmeyer, Jan, Volkmann-Kohlmeyer, Brigitte, Mostert, Lizel, O'Donnell, Kerry, Sipman, Harrie, Rogers, Jack D., Shoemaker, Robert A., Sugiyama, Junta, Summerbell, Richard C., Untereiner, Wendy, Johnston, Peter R., Stenroos, Soili, Zuccaro, Alga, Dyer, Paul S., Crittenden, Peter D., Cole, Mariette S., Hansen, Karen, Trappe, James M., Yahr, Rebecca, Lutzoni, François, & Spatafora, Joseph W. 2009a. The Ascomycota Tree of Life: A Phylum-wide Phylogeny Clarifies the Origin and Evolution of Fun-

damental Reproductive and Ecological Traits. *Systematic Biology*, **58**(2), 224–239.

Sinha, Sunita, Flibotte, Stephane, Niera, Mauricio, Formby, Sean, Plemenitaš, Ana, Gunde Cimerman, Nina, Lenassi, Metka, Gostinčar, Cene, Stajich, Jason E., & Nislow, Corey. 2017. Insight into the Recent Genome Duplication of the Halophilic Yeast Hortaea werneckii: Combining an Improved Genome with Gene Expression and Chromatin Structure. *G3: Genes, Genomes, Genetics*.

Stamatakis, Alexandros. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)*, **30**(9), 1312–3.

Sugiyama, Junta, Hosaka, Kentaro, & Suh, Sung-Oui. 2006. Early diverging Ascomycota: phylogenetic divergence and related evolutionary enigmas. *Mycologia*, **98**(6), 996–1005.

Tavares, Sílvia, Ramos, Ana Paula, Pires, Ana Sofia, Azinheira, Helena G, Caldeirinha, Patrícia, Link, Tobias, Abranches, Rita, Silva, Maria do Céu, Voegele, Ralf T, Loureiro, João, & Talhinhas, Pedro. 2014. Genome size analyses of Pucciniales reveal the largest fungal genomes. *Frontiers in plant science*, **5**, 422.

Wallace, Iain M, O'Sullivan, Orla, Higgins, Desmond G, & Notredame, Cedric. 2006. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic acids research*, **34**(6), 1692–9.

Wehe, André, Bansal, Mukul S., Burleigh, J. Gordon, & Eulenstein, Oliver. 2008. DupTree: A program for large-scale phylogenetic analyses using gene tree parsimony. *Bioinformatics*, **24**(13), 1540–1541.

Zajc, Janja, Liu, Yongfeng, Dai, Wenkui, Yang, Zhenyu, Hu, Jingzhi, Gostinčar, Cene, & Gunde-Cimerman, Nina. 2013. Genome and transcriptome sequencing of the halophilic fungus Wallemia ichthyophaga: haloadaptations present and absent. *BMC Genomics*, **14**(1), 617.

Chapter 7

Genome Sequence of the Brown Rot Fungal Pathogen *Monilinia laxa*

Naranjo-Ortiz, M. A. et al. Genome sequence of the brown rot fungal pathogen *Monilinia laxa*. Genome Announc. **6**, 1–2 (2018). URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5920163/ doi: 10.1128/genomeA.00214-18

Naranjo-Ortíz MA, Rodríguez-Píres S, Torres R, De Cal A, Usall J, Gabaldón T. Genome Sequence of the Brown Rot Fungal Pathogen Monilinia laxa. Genome Announc. 2018 Apr 26;6(17). DOI: 10.1128/genomeA.00214-18

Chapter 8

Summarizing discussion

A new hypothetical evolutionary scenario for fungal terrestrialization

One of the main novelties in this work is the proposal of a novel evolutionary scenario that aims to explain the process of fungal terrestrialization. The traditional paradigm regarding the biosphere during the Criogenian is a pessimist one. For instance, the Wikipedia article of "Snowball Earth" contains a section titled "Survival of life through frozen periods" in which the general assumption is made that life should have not been able to survive a global glaciation event. Under this assumption they start discussing that life could have survived if the glaciation were not absolute or through small microniches. We consider that this point of view is erroneous in a microbial context (it might be correct for mega-fauna and flora) and that frozen environments have the potential to have actually promoted evolutionary radiations for microbial clades.

The traditional paradigm stems from the assumption that cold is stressful for life and we consider that this is difficult to sustain in light of our knowledge of arctic, antarctic and high altitude ecosystems. Polar oceans present high biomass and a considerable diversity of macroscopic life. The taxonomic distribution of these organisms is not restrictive, which suggests that adaptation to this kind of environments is an easy evolutionary step. On top of that, many macroscopic organisms present life stages that are able to tolerate complete freezing of their tissues, including organisms that are not generally found in cold places. When we move the discussion to the microbial world we have to acknowledge that cold does indeed limit the speed of metabolic reactions. But slowing life is not the same as stopping life. Our fridges are constant reminders that bacteria and fungi can thrive in low temperatures. Most of the organisms that spoil food are airborne or present in the original aliment, and quite often present quite high optimal growth temperature. This results are in accordance to direct field observations and metagenomic studies. Snow, for instance, is commonly colonized by algae or bacteria (Anesio *et al.* , 2017). Glaciers and polar ice sheets present micro-environments where liquid water can be found, specially due to accumulation of dissolved salts that are excluded during the formation of ice crystals; or due to higher pressures (I. e. liquid layers at the bed of a glacier) and the microbiota of these microniches is extremely diverse and abundant. In conclusion, the main limitation that ice imposes to life is water limitation, rather than temperature, as solid water is not available for biological processes. At microscopic scale, though, ice surfaces might present tiny, sometimes intermittent, layers of liquid water that can be used by microbial life to proliferate. It is very important to note that ice allow for these and many other microniches, while water columns tend to form simple gradients. In fact, open ocean is a considerably boring environment, formed by a photic region that provides photosynthetic carbon

to deeper areas by precipitation, with a gradient based on light, gas composition and temperature. Continental marine environments are much more dynamic, but in most cases this is achieved by the presence of complex organisms that populate the benthos and structure the communities. Since most macroscopic community driver groups did not exist during the Cryogenian, one should not underestimate the massive ecosystem diversification that would provide a glaciation event, even though one does not need such events to have ice sheets in the planet. The diversity of microniches, the highly unstable environments, intermittent nutrient and water limitation, and the existence of a solid interface are properties that are common to both soils and ice environments. In the absence of soil macro-organisms are able to disaggregate the bedrock during soil formation, ice does act as a powerful erosive agent. Thus, ice might have been directly responsible for the formation of some of the earliest edaphic systems. Compared to sediments, ice environments present a high oxygen concentration both due to the direct exposition to the atmosphere and the incremented solubility of the gases at low temperatures. This incremented oxygen availability is an overlooked factor that should have promoted eukaryotic life in those environments, an important limiting factor in aquatic environments. From the point of view of microbes Snowball Earth was certainly not a catastrophe, but more likely a revolution.

Related to this, there is just another assumption that is inexplicably common, and that is the premise that fossil record is anything other that a tiny distorted window to life in ancient times. This premise taints the interpretation of the early evolution of not only fungi, but also metazoans and plants. Fossil evidence is powerful, but limited. In combination with indirect lines of evidence (molecular phylogenies, paleobiogeochemistry) and our knowledge of taphological processes, we should be able to pinpoint gaps in our knowledge and try to fill them by extrapolation. The transition to land environment in particular has been extremely distorted by the refusal to acknowledge well known limitations of fossilization processes. Not only that, we should start acknowledging limitation of direct observation of contemporary environments. These limitations are particularly egregious when considering microbial life. Not only microbes are extremely difficult to fossilize, but in many cases it is incredibly difficult to classify a microscopic fossil structure to a known clade. Organisms might exist in the environment as important components of the trophic chains without presenting a high enough population density to justify that "if they exist they must be represented in our data". Zoosporic fungi, for instance, are commonly parasitic in nature, and as such one would expect that their population density is high only as a response of dramatic population increases of their hosts, such as algal blooms. Failing in acknowledging these limitations of the fossil record has, in our opinion, generated extremely unrealistic paradigms to understand ecological transitions, specially the conquest of land by different groups of organisms. As an example, let us discuss for a moment the process of arthropod terrestrialization. The overall paradigm is somehow based on crabs and other semi-aquatic macroscopic arthropods. In the case of Arachnida and Miriapoda, the existence of fossils of *bona fide*, albeit chronologically late macroscopic fossils of these groups in aquatic environments has been traditionally interpreted as evidence of a clear semi-aquatic transitional stage. However, it is very naïve to assume that all prehistorical arthropods were relatively large. Indeed, a great portion of current arthropods are micro-invertebrates that live in the intersticia of soils and sediments. Terrestrialization in metazoans is not exclusive of macroscopic members of the Arthropoda, Craniata, Mollusca, Annelida, Onychophora and Platyhelminthes, as terrestrial members of the Tardigrada, Nematoda and Rotifera have also colonized soil environments. On top of that any of these three phyla is ancient enough as to have played key early roles in shaping terrestrial life. The same logic applies to protozoans, which are common inhabitants of soils. Taken together, and coming back to the subject of fungal terrestrialization, we think it should not be surprising that Zoopagomycota, the first splitting lineage of purely terrestrial fungi are associated to soil micro-invertebrates and protozoans. From the physicochemical point of view, switching from sediment to soil allow a much smoother transition from water to land that the ever popular "crawling fish". While it varies wildly, in most cases air within soil is very humid, allowing free gas exchange without much danger of dehydration. On top of that, radiation is highly mitigated by solid particles, allowing life without much photoprotection buried at mere centimeters deep in the dirt. The same can be said from snow and ice, and we considered cold desert, tundra and high elevation edaphic communities have the potential to revolutionize our understanding about terrestrialization processes.

The biggest concern for this particular hypothesis is that, so far, we cannot think in a single way to support it with direct evidence. Cold environment transitions are considerably smooth ones, and thus the genomic imprint that might have left is probably undetectable. The taxonomic composition of fungi in ice environments might reflect a phylogenetic pattern that could support the hypothesis, and in fact some of these environments such as peri-glacier soils and arctic waters present a very high biodiversity of zoosporic fungi. However, these fungi could have arrived during more recent colonizations, with the original lineages having become extinct. No study has reported the diversity of Zoopagomycota in these kind of environments, but a good support for the hypothesis would be the finding basal lineages within this phylum inhabit ice environments. Hypothetical fossil evidence of early filamentous fungi associated to glacier activity in the Cryogenian would be a good support for the hypothesis. This evidence might never be available, due to the limitations of the fossil record. All these shortcuts, however, are not unique to our proposed hypothesis. The only hypothesis that present some level of direct evidence is the "green scenario", as enzymes related to metabolism of plant cell wall compounds seem to be very ancient within the fungi (Lücking *et al.* , 2009). However, the presence of these enzymes is not direct indication of terrestrialization through direct association with Streptophyta. Fungi might have evolved from relatively specialized parasites of unicellular algae or might have been adapted to exploitation of Streptophyta derived necromass. Even more, the study does not take into account possible horizontal gene transfer across fungi. Last, there is a fourth possible scenario that we have omitted due to the early stage of the research on some of their premises. There is a growing evidence of the existence of large microbial communities in lithospheric habitats, possibly up to several kilometers deep (Ivarsson *et al.* , 2016). Fungi seem to be components of such environments. Most studies so far appear to have identified fungi from well known distal lineages, although the sheer size of this habitat still holds the possibility of finding novel lineages that might revolutionize our views in the fungal tree of life. Even if not, lithosphere might have served as intermediate stages between aquatic and dry land, from oceanic crust to continental crust, and from there to superficial land. So far, this idea falls in the terrain of scientific speculation, although the recent claims of fungal-like organisms in 2.4 Gya basalt (Bengtson *et al.* , 2017) might soon ignite the debate of the role of deep biosphere in the evolution of fungi in particular and eukaryotes in general.

Horizontal gene transfer of D-amino acid metabolic genes

We believe the implications of our study in HGT of racemases are deeper that what was discussed within the paper and that it deserves further analysis. As far as we know, our study shows several examples of serial HGT in fungi for the first time. This kind of events, in which a fungus acquires a gene from an external source, mostly by bacteria; and later acts as a donor to other fungi, have deep implications for the evolution of the metabolic capabilities of fungi, specially within the Pezizomycotina. The fact that some of these genes have been reacquired from fungi, instead of independently acquired from bacteria several times implies that the barriers for horizontal gene transfer between fungi are lower than the barriers from bacteria to fungi. In turn, this result might invalid most estimations of HGT across the fungi. A previous analysis by our group reported that approximately 1.5% of the proteome of members of the Pezizomycotina was acquired by HGT from prokaryotes, while the rates were much lower in other clades of the tree of life. HGT from other eukaryotes, and specially from other fungi might considerably expand this threshold. The decision to explore only prokaryotic genes obeys a methodological reasoning, as these genes generate a phylogenetic discordance that is much easier to identify. Fungi to fungi HGT is very difficult to distinguish from incomplete lineage sorting or differential gene loss, and thus its study is extremely difficult. For bacteria, the existence of regions with considerable composition differences has been used to pinpoint HGT, but there is evidence suggesting that this approach is invalid in eukaryotes due to the speed at which compositional signals are erased from the genome (Strese *et al.* , 2014). Thus, phylogenomic approaches are needed for the identification of such events, which in turn requires extensive taxon sampling, extensive computational resources and a considerable amount of manual labor. Population studies are able to detect genetic exchange between populations, and should be able to trace certain events under the right conditions. Unfortunately, the event would be missing if it occurred before the divergence of all the analyzed populations, thus restricting

this approach to very short evolutionary distances. Phylogeny based methods must face another problem that the field has been trying to avoid for decades: Phylogenetic assumptions are not universally valid. The phylogenetic signal for many protein families is highly volatile and might be lost even at very short evolutionary distances. Even worse, certain protein families evolve by architectural shifts, rather than point mutations. This last case even opens the possibility of "horizontal domain transfer", a phenomenon that might be much more important that previously thought in the evolution of the biosynthetic pathways of certain families of secondary compounds that are widespread in fungi, such as non ribosomal peptides and polyketides.

The ecological and physiological implications of HGT are extensive. In theory these processes, rare as they might be, hold the potential for any particular fungus to access a nearly infinite pool of metabolisms from environmental prokaryotes. In this regard it is highly plausible that HGT are far from a rarity and that the observed rates are the result of environmental selective forces purging the genes rather than of a low flow of genes between species. The exploration of fungal pangenomes hold the potential of solving these questions, but it is noteworthy that the results in *Saccharomyces cerevisiae*, one of the few species for which true pangenomes are available (Peter *et al.* , 2018), support this data. In this yeast, adaptation to the unusual niche of wine fermenting has been related not only with the formation of novel hybrid lineages, but also with the acquisition of new genes by HGT. While HGT from prokaryotes opens up a high diversity of metabolic pathways for fungi, gene flow from other fungi might provide much more interesting functionalities. For instance, transfer of effector genes is involved in host switch in several plant pathogenic fungi and it might be a much more widespread phenomenon once considered other poorly explored fungal niches, such as endophytic fungi, entomopathogens or mycoparasites. The limited interest in this last fungal niche is a great tragedy, as we are sure that these fungi are hiding many secrets that would greatly contribute to agronomy, medicine and ecology. Regarding our discussion, the role of mycoparasitism in the propagation of HGT acquired genes has barely been explored, limited to a few studies with *Parasitella parasitica* (Kellner *et al.* , 1993). However, the ubiquity of these interactions and the disruption of the sturdy cell wall caused by these pathogens holds the potential for facilitating HGT, both by direct transference from the invader hyphae and by weakening the barriers for transference from environmental microbes.

The experimental treatment of these events has an old albeit minimal history behind. The formation of heterokaryons by induction of parasexual mating in filamentous Ascomycota is far from new, and it was a powerful tool during the golden ages of fungal genetics (Strom & Bushley, 2016). The same can be said from the reports of direct conjugation between yeasts and bacteria and the formation of extremely divergent heterokaryons by protoplast fusion. All of these approaches faded into obscurity first with the development of more efficient transformation tools, and later with the rise of genomics. However, the mechanisms of HGT in fungi are still unknown, and experimental approaches are necessary to elucidate this phenomenon. Transformation by *Agrobacterium* and similar bacteria is possible in fungi, and might be happening in natural conditions in a regular basis (Lacroix & Citovsky, 2016; Suzuki *et al.* , 2015). However, as we show with our results, fungi to fungi transfer is considerably common, and thus this cannot be the only mechanism involved. It is very reasonable to assume that, under highly stressful circumstances, reproductive barriers to mating would be lowered. We could, for instance, design an experimental system in which two filamentous fungi are transformed with two different resistance markers and forced to grow through a medium containing both antifungals. If the fungi are able to grow through the selective medium, that would imply that they have either transferred the resistance genes or that they have hybridized, with the second scenario being more surprising the larger the evolutionary distance between the two is. Similar scenarios are possible in nature, specially if two antifungal producing fungi are fighting each other over a certain territory. If such scenario reaches a draw, the only logical path forward is HGT or hybridization. Once the selective pressures that hold together such improbable hybrids or that keep functioning horizontally acquired genes are removed from the environment, the fungi would probably tend to regain the original state. Analogously, old reports of conjugation between yeasts and bacteria (Heinemann & Sprague, 1989,9) should be expanded to cover a wider taxonomic range of both, as this could be a common phenomenon in natural populations.

Validation is yet another problematic aspect of HGT detection. Biological samples are often con-

taminated with different sources of DNA, including microbes from the environment and associated to the sample itself, as well as human-caused contamination during handling. These extraneous sequences are sometimes difficult to distinguish from *bona fide* HGT events. The opposite holds true, and it is likely than many true HGT events have been filtered out by automatic pipelines designed to get rid of contamination. Contamination is rather easy to identify if the problem genome is well assembled, as a true HGT event should be surrounded by genes whose phylogenetic signal matches the organism's natural history. In this regard, PacBio and Oxford Nanopore technologies should enhance our confidence in HGT claims by allowing higher quality genome assembly. Transcriptomic evidence is another excellent approach to validate HGT events, although certain circumstances such as microbes associated with the sample might confound the results. The presence of the putative HGT acquired genes in several phylogenetically related organisms might be a good indication that the event is genuine, as one would expect that contamination would be an accidental event rather than a systematic error.

Karyon:

As the field of genomics ventures farther and farther away, we are expecting to find new exotic genomic architectures that would defy our imagination and methodological approaches. Karyon was developed with this paradigm in mind. We have extensively reviewed the kind of unorthodox genomes we have expected and we have theorized that many others might exist in nature. However, we consider that the applicability of our approach goes far beyond of fungi and that the protistology and zoology community are likely to benefit from Karyon or similar methods that are yet to arrive. Hybrids are well known for animals, and basically any clade of eukaryotes with a sexual cycle is expected to present hybrids or, at least, highly heterogeneous genomes. Oomycetes share mycelial growth with fungi, and as such they might be susceptible to develop heterokaryotic mycelia. A similar situation could be encountered in social amoebae, many of which form plasmodia by accretion of many independent single-celled stages. In the presence of mixed environmental populations, plasmodia might as well ended up formed by genetically heterogeneous nuclei. This type of systems were very popular for theoretical biologists during the second half of the XXth century and we would be very excited for the return of this "retro fashion", now that we have genomic technologies to complement it. Aneuploidies are also very common in protozoans, at least for parasitic lineages such as *Giardia* or the trypanosomids (Berman, 2016). For these groups it has been described the existence of meta-stable mosaic aneuploid populations, in which aneuploid appear in a constant proportion of the population. The exact proportions and particular chromosomes affected are dependent on the conditions, and these populations emerge again and again independently, regardless of the characteristics of the initial population. Polyploidy is also relatively common in certain lineages, and some spectacular cases have already been described. Several species of the plant parasitic nematode genus *Meloidogyne* are asexual and seem to present their genome in several copies. In some cases, the chromosomes of these nematodes seem to have suffered extensive recombination to the point that most of them are unable to recombine as a whole with any other chromosome (Blanc-Mathieu *et al.* , 2017). This situation implies that any particular region is present in a certain, more or less narrow range of copies; but that the co-linear sequences to that particular region vary from copy to copy. With the exception of ciliates, we are confident that these nematodes present the "craziest" genome architecture we have ever heard about.

Due to the necessity of an accurate variant calling process, Karyon does require high quality reads coming from Illumina technologies. These technologies have crushed any competence for many years due to the raw output volume, low prize and high per base quality; and as such we are confident that Karyon should be applicable to most past and current projects. The undisputed reign of Illumina is expected to end some day in a not so distant future, with novel sequencing platforms such as PacBio or Oxford Nanopore promising an alternative with many interesting properties for genome assembly. The reads produced by these novel technologies are extremely long and should help to solve many limitations of current approaches. This includes the resolution of highly repetitive regions or the identification of certain chromosomic rearrangements. The

limitation, however, stems from the high error rate that both of these methods present, which hold the potential of introducing many mistakes during variant calling protocols. The community is still developing novel assembly software that are able to fully employ the promising potential of these technologies. In this regard, we consider that it is important to start breaking old paradigms on genomics that are based on extrapolations from a handful of model organisms. For instance, heterozygosity is a common feature of most genomes, but we have traditionally cherry-picked those genomes in which this value is very low. Heterozygosity does in act contain intrinsic information and some novel tools, such as BWISE (https://github.com/Malfoy/BWISE), aim to use it as an advantage rather than a limitation.

1KFG phylogenomic analyses

Duplication rate analyses While the analyzed dataset for this project is truly massive, the representation of different taxa is fairly irregular. Representation is rather poor for the zoosporic lineages, as well as for Zoopagomycota, Glomeromycota and Mucoromycota. Nothing can be said about he zoosporic lineages, as the sampling remains too sparse and no WGD event has ever been described in any of these lineages. Regarding Zoopagomycota, our data shows a rather small peak within Kickxellomycotina, as well as another the genus *Basidiobolus*. The robustness of these claims is compounded by the lack of representation in the clade. A similar situation occurs for *Mortierella* (Mortierellomycotina), *Rhizopus microsporus* (Mucoromycota, with an already described WGD event (Corrochano *et al.* , 2016; Ma *et al.* , 2009)), the two represented species of Neocallimastigomycotina and *Allomyces* (Blastocladiomycota), where our analyses detect high duplication frequency. In addition to that, data from a collaborative work shows a probable hybridization event in *Condidiobolus*. Mucoromycota is poorly represented in our particular dataset, but not so much in general databases. As a consequence, Mucorales have several described WGD events that involve *Mucor* and *Rhizopus*, the two best sampled genera. In addition to that, another collaborative work, as well as our survey reveals the existence of WGD in the genus *Lichtheimia*, as well as another probable event in *Mucor racemosus*, previously undescribed. If we add our results in Basidiomycota and the well known cases of hybridization and WGD within the Saccharomycotina, we obtain a pretty clear picture. Hybridization and WGD is common in fungi, as events keep popping by different methodologies as soon as the genomic coverage of a given taxa reaches a certain minimum. WGD reports are not very common in Pucciniomycotina and Ustilaginomycotina, although in these cases it might be due to reiterative biases toward certain well known groups, combined with an still limited genomic representation. Taphrinomycotina, on the other hand, presents highly reduced genomes and a very reduced number of described lineages. However, the main out-layer in this trend is, by far, Pezizomycotina.

Pezizomycotina represents the most diverse lineage of fungi, and the representation in databases, albeit biased in favor of certain groups, is truly impressive. Yet, duplication rates remain consistently low for all sampled lineages so far, with only the extremotolerant *Hortaea werneckii* showing signs of WGD (Lenassi *et al.* , 2013). Surprisingly, the proteome size of filamentous Pezizomycotina can be highly variable even between species of the same genera. Part of this dissonance might be due to HGT. Transferred genes with a very distant donor would probably appear as orphan genes, while at smaller phylogenetic distances the phenomenon has the possibility of greatly disrupting automated tree topology analysis algorithms. Beyond hypothetical artefactual analytic results, higher HGT rate might imply that gene duplication is relatively unimportant for these fungi as a source of genetic and metabolic innovation. There might exist ecological constrains to genome size that are preventing the survival of superfluous duplicated genes at long evolutionary distances. However, such constrains seem to exist in some plant lineages, and this has not prevented the detection of high duplication rates interpreted as paleopolyploidization in the highly reduced genomes of *Arabidopsis* and *Utricularia* (Jiao *et al.* , 2011; Soltis *et al.* , 2015). Pezizomycotina, unlike Basidiomycota, do not usually present controlled nuclear divisions, which allow for all the implications exposed in chapter 3. *A priori*, it is an stretch to relate this possibility to all the peculiarities of Pezizomycotina, such as low duplication rates, mesosynteny, high HGT rates and so on; but we consider that the topic deserves further exploration. If, for other species, duplication

of genes obey relatively punctual circumstances that are afterwards fixed in the genome of populations, the ability to form functional heterokaryons with the properties previously exposed could help to explain the numbers. The association of genes in clusters, mesosynteny included, have been proposed to be caused by HGT (Wisecaver & Rokas, 2015). For the sake of the transference, the order of the genes is less important than the completion of the biochemical pathway, and thus physical linkage of the genes would help HGT. Phylogenetic evidence offer little support for long range HGT affecting cluster formation. However, the formation of heterokaryons could potentially allow repeated transference of such clusters among closely related strains by mean of translocations and similar chromosomal rearrangements between pairs in parasexual processes. As this exchange would mostly occur at the level of populations or relatively closely related species, phylogenomic approaches would lack resolution to explore this kind of processes.

Conclusions

1. We propose a new hypothetical evolutionary scenario for fungal terrestrialization, the "white scenario". Under this scenario aquatic zoosporic fungi would have acquired adaptations required for life on dry land through adaptation to frozen environments. We review evidence supporting this hypothesis.

2. We propose an ecological definition of the term "yeast" that can be summarized roughly as a non-territorial or prokaryote-like fungus. We fungal lineages that share a series of ecological, biochemical and genomic characteristics that suggest convergent evolution towards this lifestyle. We provide arguments to move away from the current use of the term "yeast", which we consider highly confusing and based on arbitrary phenotypical characterization:

3. We provide compelling evidence of the existence of serialized horizontal gene transfer events in fungi. We interpret this finding as evidence that fungal species have access to a wide environmental gene pool mostly through the exchange of genes with other fungi, not unlike many studied prokaryotes.

4. Genes related to D-amino acid metabolism take part on horizontal gene transfer events at high frequency. Throughout the eukaryotic tree of life, we have described a total of 48 novel HGT events involving these genes, which accounts for approximately 7.0% of the genes of this kind in eukaryotes. These events seem to be particularly common in plant pathogenic fungi, for which we propose a role in nitrogen acquisition when growing in plant tissues.

5. We present Karyon, an automated pipeline for the assembly and genome architecture analysis of problematic genomes. We have applied said pipeline to a collection of 32 poorly assembled genomes from public databases in the Mucorales, from which five appear to present some kind of genomic anomaly. This suggests non-canonical genomic architectures might be common in fungi.

6. We obtained high phylogenetic support for a clade formed by Wallemiomycetes and Agaricomycotina, within Basidiomycota,

7. Our re-evaluation of the phylogenetic relationships between the different groups of Taphrinomycetes recover sisterhood of Archaeorhizomycetes with Neolectomycetes; as well as sisterhood of Pneumocystidomycetes with Schizosaccharomycetes. Taphrinomycetes, however, appears paraphyletic in our analyses, which raising doubts about the placement of *Saitoella*.

8. Phylogenomic analyses show that large-scale gene duplication events are extremely common across all fungal groups except Ascomycota. Most of the duplication peaks appear in terminal nodes, but in at least one case an internal duplication peak can be related to the acquisition of lignin degrading capabilities and the diversification of body plan in Agaricomycetes.

9. The present thesis reviews current knowledge regarding the role of non-vertical genomic events in the evolution of the Fungi. We have developed novel methodologies for the study of these phenomena and applied it to diverse clades of Fungi at different evolutionary scales. We conclude that non-vertical genomic phenomena are very common in Fungi and act as powerful evolutionary drivers across virtually all scales and clades within the fungal tree of life.

References

Anesio, Alexandre M., Lutz, Stefanie, Chrismas, Nathan A. M., & Benning, Liane G. 2017. The microbiome of glaciers and ice sheets. *npj Biofilms and Microbiomes*, **3**(1), 10.

Bengtson, Stefan, Rasmussen, Birger, Ivarsson, Magnus, Muhling, Janet, Broman, Curt, Marone, Federica, Stampanoni, Marco, & Bekker, Andrey. 2017. Fungus-like mycelial fossils in 2.4-billion-year-old vesicular basalt. *Nature Ecology & Evolution*, **1**(apr), 0141.

Berman, Judith. 2016. Ploidy plasticity: a rapid and reversible strategy for adaptation to stress. *FEMS Yeast Research*, **16**(3), fow020.

Blanc-Mathieu, Romain, Perfus-Barbeoch, Laetitia, Aury, Jean-Marc, Da Rocha, Martine, Gouzy, Jérôme, Sallet, Erika, Martin-Jimenez, Cristina, Bailly-Bechet, Marc, Castagnone-Sereno, Philippe, Flot, Jean-François, Kozlowski, Djampa K., Cazareth, Julie, Couloux, Arnaud, Da Silva, Corinne, Guy, Julie, Kim-Jo, Yu-Jin, Rancurel, Corinne, Schiex, Thomas, Abad, Pierre, Wincker, Patrick, & Danchin, Etienne G. J. 2017. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLOS Genetics*, **13**(6), e1006777.

Corrochano, Luis M, Kuo, Alan, Marcet-Houben, Marina, Polaino, Silvia, Salamov, Asaf, Villalobos-Escobedo, José M, Grimwood, Jane, Álvarez, M Isabel, Avalos, Javier, Bauer, Diane, Benito, Ernesto P, Benoit, Isabelle, Burger, Gertraud, Camino, Lola P, Cánovas, David, Cerdá-Olmedo, Enrique, Cheng, Jan-Fang, Domínguez, Angel, Eliáš, Marek, Eslava, Arturo P, Glaser, Fabian, Gutiérrez, Gabriel, Heitman, Joseph, Henrissat, Bernard, Iturriaga, Enrique A, Lang, B Franz, Lavín, José L, Lee, Soo Chan, Li, Wenjun, Lindquist, Erika, López-García, Sergio, Luque, Eva M, Marcos, Ana T, Martin, Joel, McCluskey, Kevin, Medina, Humberto R, Miralles-Durán, Alejandro, Miyazaki, Atsushi, Muñoz-Torres, Elisa, Oguiza, José A, Ohm, Robin A, Olmedo, María, Orejas, Margarita, Ortiz-Castellanos, Lucila, Pisabarro, Antonio G, Rodríguez-Romero, Julio, Ruiz-Herrera, José, Ruiz-Vázquez, Rosa, Sanz, Catalina, Schackwitz, Wendy, Shahriari, Mahdi, Shelest, Ekaterina, Silva-Franco, Fátima, Soanes, Darren, Syed, Khajamohiddin, Tagua, Víctor G, Talbot, Nicholas J, Thon, Michael R, Tice, Hope, de Vries, Ronald P, Wiebenga, Ad, Yadav, Jagjit S, Braun, Edward L, Baker, Scott E, Garre, Victoriano, Schmutz, Jeremy, Horwitz, Benjamin A, Torres-Martínez, Santiago, Idnurm, Alexander, Herrera-Estrella, Alfredo, Gabaldón, Toni, & Grigoriev, Igor V. 2016. Expansion of Signal Transduction Pathways in Fungi by Extensive Genome Duplication. *Current biology*, **26**(12), 1577–1584.

Heinemann, Jack A., & Sprague, George F. 1989. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature*, **340**(6230), 205–209.

Heinemann, Jack A., & Sprague, George F. 1991. Transmission of Plasmid DNA to Yeast by Conjugation with Bacteria. *Methods in Enzymology*, **194**(C), 187–195.

Ivarsson, M., Bengtson, S., & Neubeck, A. 2016. The igneous oceanic crust – Earth's largest fungal habitat? *Fungal Ecology*, **20**(apr), 249–255.

Jiao, Yuannian, Wickett, Norman J, Ayyampalayam, Saravanaraj, Chanderbali, André S, Landherr, Lena, Ralph, Paula E, Tomsho, Lynn P, Hu, Yi, Liang, Haiying, Soltis, Pamela S, Soltis, Douglas E, Clifton, Sandra W, Schlarbaum, Scott E, Schuster, Stephan C, Ma, Hong, Leebens-Mack, Jim, & DePamphilis, Claude W. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature*, **473**(7345), 97–100.

Kellner, Martina, Burmester, Anke, Wöstemeyer, Anke, & Wöstemeyer, Johannes. 1993. Transfer of genetic information from the mycoparasite Parasitella parasitica to its host Absidia glauca. *Current Genetics*, **23**(4), 334–337.

Lacroix, Benoît, & Citovsky, Vitaly. 2016. Transfer of DNA from Bacteria to Eukaryotes. *mBio*, **7**(4), e00863–16.

Lenassi, Metka, Gostinčar, Cene, Jackman, Shaun, Turk, Martina, Sadowski, Ivan, Nislow, Corey, Jones, Steven, Birol, Inanc, Cimerman, Nina Gunde, & Plemenitaš, Ana. 2013. Whole Genome Duplication and Enrichment of Metal Cation Transporters Revealed by De Novo Genome Sequencing of Extremely Halotolerant Black Yeast Hortaea werneckii. *PLoS ONE*, **8**(8), e71328.

Lücking, Robert, Huhndorf, Sabine, Pfister, Donald H., Plata, Eimy Rivas, & Lumbsch, H. Thorsten. 2009. Fungi evolved right on track. *Mycologia*, **101**(6), 810–822.

Ma, Li-Jun, Ibrahim, Ashraf S, Skory, Christopher, Grabherr, Manfred G, Burger, Gertraud, Butler, Margi, Elias, Marek, Idnurm, Alexander, Lang, B Franz, Sone, Teruo, Abe, Ayumi, Calvo, Sarah E, Corrochano, Luis M, Engels, Reinhard, Fu, Jianmin, Hansberg, Wilhelm, Kim, Jung-Mi, Kodira, Chinnappa D, Koehrsen, Michael J, Liu, Bo, Miranda-Saavedra, Diego, O'Leary, Sinead, Ortiz-Castellanos, Lucila, Poulter, Russell, Rodriguez-Romero, Julio, Ruiz-Herrera, José, Shen, Yao-

Qing, Zeng, Qiandong, Galagan, James, Birren, Bruce W, Cuomo, Christina A, & Wickes, Brian L. 2009. Genomic analysis of the basal lineage fungus Rhizopus oryzae reveals a whole-genome duplication. *PLoS genetics*, **5**(7), e1000549.

Peter, Jackson, De Chiara, Matteo, Friedrich, Anne, Yue, Jia-Xing, Pflieger, David, Bergström, Anders, Sigwalt, Anastasie, Barre, Benjamin, Freel, Kelle, Llored, Agnès, Cruaud, Corinne, Labadie, Karine, Aury, Jean-Marc, Istace, Benjamin, Lebrigand, Kevin, Barbry, Pascal, Engelen, Stefan, Lemainque, Arnaud, Wincker, Patrick, Liti, Gianni, & Schacherer, Joseph. 2018. Genome evolution across 1,011 Saccharomyces cerevisiae isolates Species-wide genetic and phenotypic diversity. *Nature*, **556**.

Soltis, Pamela S, Marchant, D Blaine, Van de Peer, Yves, & Soltis, Douglas E. 2015. Polyploidy and genome evolution in plants. *Current Opinion in Genetics & Development*, **35**(dec), 119–125.

Strese, Ake, Backlund, Anders, & Alsmark, Cecilia. 2014. A recently transferred cluster of bacterial genes in Trichomonas vaginalis–lateral gene transfer and the fate of acquired genes. *BMC evolutionary biology*, **14**(1), 119.

Strom, Noah B, & Bushley, Kathryn E. 2016. Two genomes are better than one: history, genetics, and biotechnological applications of fungal heterokaryons. *Fungal biology and biotechnology*, **3**, 4.

Suzuki, Katsunori, Moriguchi, Kazuki, & Yamamoto, Shinji. 2015. Horizontal DNA transfer from bacteria to eukaryotes and a lesson from experimental transfers. *Research in Microbiologoy*, **166**(10), 753–763.

Wisecaver, Jennifer H., & Rokas, Antonis. 2015. Fungal metabolic gene clusters: caravans traveling across genomes and environments. *Frontiers in Microbiology*, **6**(mar), 161.