UNIVERSITY OF THESSALY

DEPARTMENT OF BIOCHEMISTRY AND BIOTECHNOLOGY

POSTGRADUATE PROGRAM

APPLICATIONS OF MOLECULAR BIOLOGY – MOLECULAR GENETICS – DIAGNOSTIC BIOMARKERS

PIPLIDOU STYLLIANI

Phylogenomics, core and accessory genomes of the *Ascomycetes* with Bioinformatics methods

LARISSA 2024

This thesis was completed in the bioinformatics laboratory of the Department of Biochemistry and Biotechnology of the University of Thessaly.

Supervisor:

Amoutzias Grigorios: Professor of Bioinformatics with emphasis in Microbiology, Department of Biochemistry and Biotechnology, University of Thessaly, Greece

Committee Members:

Amoutzias Grigorios: Professor of Bioinformatics with emphasis in Microbiology, Department of Biochemistry and Biotechnology, University of Thessaly, Greece

Oliver Stephen: Emeritus Professor of Systems Biology & Biochemistry, Department of Biochemistry, University of Cambridge, United Kingdom

Dimitris Mossialos: Associate Professor of Microbial Biotechnology, Department of Biochemistry and Biotechnology, University of Thessaly, Greece

1. Abstract

Fungi are organisms that form a distinct evolutionary group within the eukaryotes. The lineage of *Ascomycota* exhibits great diversity and encompasses numerous species that are beneficial or pathogenic. Within soil ecosystems, *Ascomycota* play a crucial role in the process of nutrient recycling. The progress of genome sequencing has made it possible to get high-quality genome assemblies of numerous fungal species. Hence, extensive comparative genomics studies are nowadays possible and may offer new insights and of great resolution on the evolution of fungal lineages and on the genes that are associated with them. In this study, we analyzed genomic data from 91 *Ascomycetes* and 2 outgroup taxa. We reconstructed a comprehensive phylogeny based on hundreds of orthologous genes/proteins and identified the core proteins of the genus, as well as the core and fingerprint proteins of the three major subgroups (*Saccharomycotina*, *Pezizomycotina*, *Taphrinomycotina*) within the *Ascomycota*.

Keywords

Bioinformatics, Phylogenomics, Core Genome, Accessory Genome, Fingerprint Proteins, Ascomycetes

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

2. Introduction

2.1. The kingdom of fungi

The fungi are one of the six kingdoms of life according to Cavalier-Smith's classification, sister to the animal kingdom (McCarthy and Fitzpatrick, 2017). They are highly diverse (Fig. 1) and they are one of the most diverse branches in the tree of life, encompassing a range of 2.2–3.8 million species. Fungi are everywhere, exerting a profound influence on the world. They have a wide range of eating habits, physical structures, growth patterns and ecological interactions, and are believed to have co-evolved with plants (Gautam et al., 2022). Fungi are frequently overlooked in ecology, however their impact on the environment and their significance in biology are unquestionable (Buckley, 2008). They play a significant role in global bio-geochemistry by efficiently recycling carbon and facilitating the movement of nitrogen, phosphorus, and other essential bio-elements.

Figure 1. Diversity of major fungal lineages (Li et al., 2021).

Fungi engage in a wide variety of symbiotic associations with animals, plants, insects, amoeba, bacteria, and other fungi (Naranjo-Ortiz and Gabaldón, 2019). Evidence continues to grow, supporting the notion that fungi play a crucial role in the composition of the gut microbiota. Fungi within the gut microbiota have been shown to impact the immune responses of the mammalian host by either reducing or enhancing local inflammatory reactions (Pérez, 2021). In addition, fungi offer crucial assistance to plant life through the presence of endophytes and mycorrhizae (Naranjo-Ortiz and Gabaldón, 2019). Approximately 60% of plants engage in symbiotic interactions with fungi; this confers advantages to the plant. This arrangement frequently gives the plant a competitive edge over plants without mycorrhizal fungi, while also providing the fungus with essential carbon. Mycorrhizal fungi seem to have played a crucial role in enabling plants to inhabit land. Bacteria also establish symbiotic relationships with fungi. Some fungal plant diseases harbor endosymbiotic bacteria that control or facilitate the synthesis of virulence components,

enabling the fungus to infect important crop plants such as rice. Additionally, these bacteria allow the fungus to generate spores throughout its life cycle, aiding in their dispersal and survival (Buckley, 2008).

Furthermore, fungi and the substances they produce are of great economic significance. Fungi have the ability to generate fermented food such as beer, wine, bread, and cheese. They also produce various primary and secondary metabolites, including ethanol, hydrolases, oxidoreductases, organic acids, vitamins, hypocholesterolemics, and antineoplastics. Fungi can be utilized as inoculants and biocides for mycorrhization and as agents for biological control. Fungi can also be employed for the purpose of bioremediation of solid waste, wastewater, and air pollutants, as well as for the production of novel biomaterials, such as mycocomposites and nanomycomaterials (Tiwary, 2020).

2.1.1. Taxonomy of fungi

The field of fungal taxonomy has gone through significant revisions, since the initial inclusion of fungi in Linnean taxonomy. The kingdom of Fungi, along with its defining characteristics and the taxa it encompasses, has been in a continual state of flux (Fig. 2) (Richards et al., 2017).

Fungal taxonomy is a major field of study that seeks to identify and categorize all fungi and their relationships. Only a small fraction, around 5%, of the estimated 2.2–3.8 million species worldwide are currently identified, resulting in a highly incomplete reconstruction of the Fungal Tree of Life (FTOL) (Zhou and May, 2023).

Currently, our understanding of the evolutionary relationships between different groups of fungi is limited since there is not enough fossil evidence available. The utilization of molecular techniques has facilitated modifications in the existing classification of fungi, providing fresh perspectives on the evolutionary aspects of fungi (Lin et al., 2023). Thanks to the advancements in new technology, mycology is entering a new era characterized by interdisciplinary collaboration and worldwide integration (Zhou and May, 2023). In the past decade, the application of advanced methods for isolating and identifying fungi has led to the establishment of several new taxa, including new divisions, classes, orders, and families (Gautam et al., 2022).

The fungal kingdom consists of a minimum of eleven distinct groups, including seven phyla and four subphyla of the polyphyletic *Zygomycota*. These taxa exhibit a

wide range of genomes, morphologies, and life cycles (Buckley, 2008). The study conducted by Hibbett et al. (2007) identified the phyla that are classified within the kingdom Fungi as follows:

1. *Ascomycota*
2. *Basidiomycota*
3. *Zygomycota*
4. *Chytridiomycota*
5. *Glomeromycota*
6. *Blastocladiomycota*
7. *Neocallimastigomycota*
8. *Microsporidia*

The *Zygomycota* is considered polyphyletic, indicating that the members of this group do not all have a common direct ancestor. Consequently, the authors have assigned the following as subphyla: The taxonomic groups *Zoopagomycotina*, *Entomophthoromycotina*, Mucoromycotina, and *Kickxellomycotina* (Buckley, 2008).

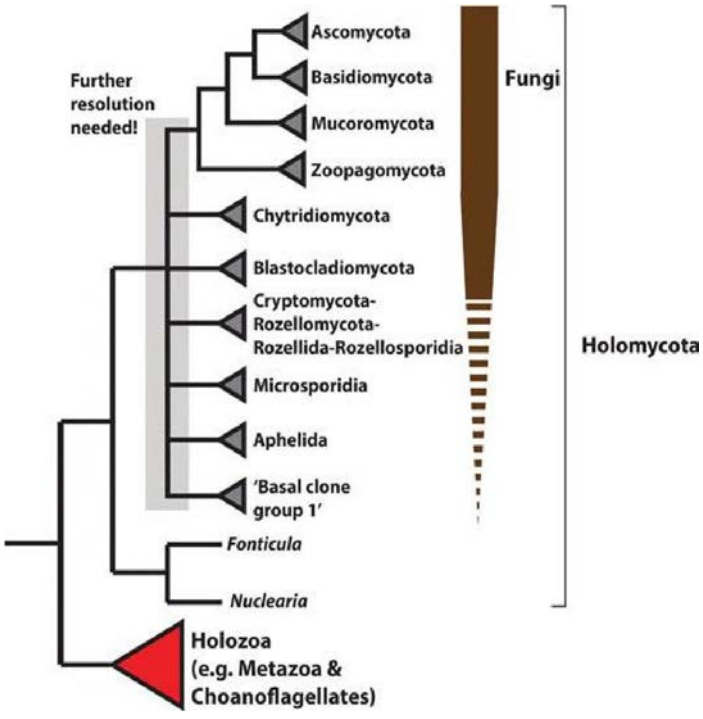


Figure 2. Schematic phylogenetic tree illustrating the major groups and the phylogenetic uncertainty among the deep branches of the Fungi (Richards et al., 2017).

Fungal species can exhibit polymorphism throughout their whole life cycle, leading to difficulties in accurately and precisely identifying them. Fungal species identification has been extensively conducted using a range of diagnostic approaches, such as phenotypic analysis, physiological profiling, and sequence-based identification (Lin et al., 2023).

The majority of the variation in terms of documented species is primarily found within the subkingdom Dikarya, which consists of *Ascomycota* and *Basidiomycota*. *Mucoromycota* is a group of organisms that are primarily connected with plants and is closely related to the Dikarya. *Zoopagomycota* and *Blastocladiomycota* are separate clades that represent deeper phylogenetic relationships among non-Dikarya fungi.

Holomycota is the most basic group that encompasses the kingdom of Fungi and is closely related to the Holozoa group (James et al., 2020).

The evolutionary links among various taxonomic ranks within the fungal kingdom are currently not sufficiently understood. An accurately determined phylogeny of fungi is essential for comprehending the evolutionary development of their genes, pathways, characteristics, and overall biology.

2.1.2. Key fungal species

Saccharomyces cerevisiae (S. cerevisiae)

Saccharomyces cerevisiae is a very popular eukaryotic model organism. It has been used as a model for investigating various biological processes, including aging, gene expression, signal transduction, cell cycle regulation, metabolism, apoptosis, neurodegenerative disorders and numerous other biological processes. For instance, it is possible that around 30% of genes associated with human diseases have corresponding genes in the yeast proteome (Karathia et al., 2011). *S. cerevisiae* is a unicellular fungus with a nuclear genomic DNA of approximately 12 Mb that is organized in 16 chromosomes (fungi.ensembl.org). Its genome was fully sequenced in 1996, in one of the foundational papers in the field of genomics and molecular biology that had the title "Life with 6000 genes" (Goffeau et al., 1996) and there are 6,600 annotated protein coding genes. The Saccharomyces Genome Database is the community resource for the budding yeast *Saccharomyces cerevisiae* and it provides information about its genome, genes, proteins, and other encoded features. *S. cerevisiae* has played a crucial role in human civilization due to its widespread application in the fermentation of food and beverages, making it commercially valuable (Parapouli et al., 2020).

Schizosaccharomyces pombe (S. pombe)

The fission yeast *Schizosaccharomyces pombe* is a unicellular eukaryote that has a rod-shaped structure. It is widely recognized for its significant role as a model organism in studying the regulation and conservation of the eukaryotic cell cycle. *S. pombe* has greater similarity to complex eukaryotes compared to the budding yeast *S. cerevisiae*. Although they are both unicellular eukaryotes and they belong to the ascomycete group, they diverged approximately 350-400 million years ago (Vyas et al., 2021). The fission yeast has a genome size of approximately 14.1Mb, organized in 3

chromosomes and it has 5,145 protein coding genes (fungi.ensembl.org). Among them, 3,536 genes have homologs in humans, and 1,244 genes are associated with diseases. *S. pombe* is abundantly found in a variety of natural sources including fruits, syrups, and a fermented tea called "Kombucha" that is made using yeasts and bacteria (Vyas et al., 2021).

Komagataella phaffii (K. phaffii)

Komagataella phaffii (*K. phaffii*; previously described as *Pichia pastoris*) is a methylotrophic yeast that belongs to the taxonomic family of *Ascomycetes*. It typically occurs in the haploid state, but nitrogen limitation stimulates mating and leads to the formation of diploid cells (Batt, 2014). *K. phaffii* has a genome size of approximately 9.4 Mb, organized in 4 chromosomes and it has 5,035 protein coding genes (fungi.ensembl.org). *K. phaffii* is significant in the industrial manufacturing of proteins. *K. phaffii* is a prominent host for the production of various heterologous proteins, satisfying industrial standards (Barone et al., 2023).

Candida albicans (C. albicans)

Candida albicans is a human opportunist pathogen. It can grow as yeast, pseudohyphae or real hyphae, both in vitro and in vivo, depending on its environment. *Candida* species collectively cause up to 400,000 instances of systemic fungal illness, resulting in mortality rates of up to 40% (Mukaremera et al., 2017). *Candida albicans* is the most common human fungal pathogen (Kim and Sudbery, 2011) and is capable of inducing two main types of infections in humans: superficial infections (such as oral or vaginal candidiasis) and serious systemic diseases that pose a threat to life (Mayer et al., 2013). *C. albicans* has a genome size of approximately 16Mb, organized in 8 chromosome pairs and it has 6,221 protein coding genes (fungi.ensembl.org).

Aspergillus nidulans (A. nidulans)

Aspergillus nidulans is an ascomycete fungus that has been commonly used in the study of significant biological phenomena such as metabolic regulation, development, differentiation, mitosis, cell cycle control and secondary metabolism. *A. nidulans* possesses unique biological characteristics that have facilitated the advancement of experimental methodologies for genome manipulation, leading to valuable experimental outcomes (Oakley, 2013). *A. nidulans* has a genome size of

approximately 30 Mbp, organized in 8 chromosomes and and it has 10,534 protein coding genes (fungi.ensembl.org).

2.2. Evolution of fungi

Cavalier-Smith has proposed a hypothesis explaining the evolution of fungi from a protist progenitor. This hypothesis suggests that fungi lost the ability to consume food particles (phagotrophy) and instead developed a method of absorbing nutrients (osmotrophy) while also developing a cell wall, frequently in the form of polarized hyphal growth (Richards et al., 2017). Although the fossil record of fungi is limited because of their basic structure, researchers have discovered fungal fossils dating back to around 400 million years ago during the Ordovician period. Molecular clock analyses indicate that fungi likely originated in the Precambrian eon, approximately 0.76–1.06 billion years ago. The integration of DNA data into phylogenetic analyses has significantly enhanced our comprehension of fungal evolution, after the implementation of phylogenetic techniques in systematics (McCarthy and Fitzpatrick, 2017).

Evolutionary novelties refer to physiological or morphological traits that are not present in ancestral lineages but are gained during the process of evolution. These traits provide an advantage for adapting to different settings and colonizing new habitats (Moriyama and Koshiba-Takeuchi, 2018). The genome of the last universal common ancestor (LUCA) consisted of approximately 500 genes. The proliferation of living forms since the LUCA has been driven by the emergence of novel genetic material through the process of evolution. Novel genes can arise through the process of de novo evolution from noncoding DNA, by gene fusions that merge preexisting domains and by gene duplication and divergence (Copley, 2020). Numerous genes exhibit patterns across genomes and taxa that cannot be explained by vertical genetic transmission or gene loss. Occasionally, they have compositional characteristics that are typical of more distantly related organisms. Therefore, the simplest explanation suggests that there is a transfer of genetic material between various strains, species, and even kingdoms (Barreiro et al., 2019).

2.2.1. Gene Duplication

Genomes are very dynamic, as a result of several ongoing processes that enable the generation of genetic novelty, which is crucial for species to undergo evolutionary changes and adapt to shifting environments. Gene duplication is a significant mechanism that offers new genetic material and the potential to acquire novel functionality (Lallemand et al., 2020). Gene duplication has played a significant role in the evolution of developmental processes across different animals, leading to the emergence of numerous new gene functions. Gene duplication can occur through mechanisms such as uneven crossing over, retroposition, or chromosomal (or genome) duplication (Magadum et al., 2013). Duplicated genes, also known as paralogs, are genes that have a homologous link due to a duplication event in the genome of a common ancestor, as opposed to a speciation event (Lallemand et al., 2020). Since the beginning of life, there has been a proliferation of new metabolic and regulatory enzymes through the process of gene duplication and divergence (Copley, 2020). In 2007, Wapinski et al. demonstrated that the process of gene duplication and loss is strongly limited by the functional characteristics and the genes' interacting counterparts. Specifically, genes associated with stress display several instances of duplication and loss, while genes related to growth have a tendency to resist such alterations (Wapinski et al., 2007).

Duplicated genes have the ability to act as a backup and effectively compensate for the absence of their duplicated counterparts. This hypothesis was validated through genome-wide gene knockout or knockdown experiments conducted in yeast and worms, which demonstrated that the effects of knockdown/knockout of duplicate genes are considerably lower than those of single genes. In contrast, studies conducted on mice in 2009 by Makino et al., found no substantial distinction in essentiality between duplicated genes and singletons. This unexpected outcome suggested that duplicate genes in mammals do not serve as a backup mechanism and highlighted the distinct mechanisms that influence the evolution and preservation of duplicate genes in mammals compared to simpler eukaryotes (Makino et al., 2009).

2.2.2. Whole Genome Duplication (WGD)

Whole genome duplications (WGDs) had a major impact upon the evolutionary process of eukaryotes. *S. cerevisiae* emerged from a whole duplication event that

occurred ~100 mya, of eight original chromosomes, and later regained its normal ploidy through losing about 90% of the duplicated genes in small deletions (Kellis et al., 2004). WGD is particularly prevalent in the plant kingdom. Additionally, there were two instances of WGD in vertebrates (the 2R event), that took place before fish and tetrapods diverged, at least 450 million years ago (Copley, 2020).

Due to the growing availability of genetic raw materials, WGD has been widely recognized as a significant factor in the development of new evolutionary traits. WGD provides a large amount of genetic material for processes such as mutation, genetic drift, and natural selection to act upon. This concept was first proposed by Ohno in 1970 (Moriyama and Koshiba-Takeuchi, 2018). Paralogs that have arisen as a result of WGD are called ohnologs (Lallemand et al., 2020).

Following WGD, duplicated genes have distinct destinies (Fig. 3). The most probable outcome of WGD is that of non-functionalization. This occurs due to the buildup of deleterious mutations in coding regions, as there is no selective pressure to preserve functionality. Eventually, this non-functionalization causes the gene to transform into a pseudogene, hence the term pseudogenization.

Subfunctionalization refers to the phenomenon in which duplicated genes have the same expression patterns or functions that initially existed in the ancestral gene. Genes often exhibit several expressions or functions, particularly in the case of transcription factors. This implies that duplicated genes have the ability to mutually share and compensate for alterations in their expressions, which can be influenced by mutations in their regulatory sequences, and in their functions, which can be influenced by changes in their coding sequences. Duplicates have the ability to be altered as long as one of them maintains their original expression or function. Subfunctionalization can arise as a result of the relaxation of selective pressures or restrictions following the duplication of the ancestral gene.

Neofunctionalization is the final result for duplicated paralogues, where one of the paralogues obtains a novel function. Neofunctionalization takes place following duplication, when the selective pressure decreases and mutations that are advantageous accumulate in regulatory or coding regions in one of the paralogues, while the other paralogue maintains its original function. Beneficial mutations are uncommon in comparison to deleterious mutations, hence neofunctionalization is also anticipated to occur less frequently following WGD than non-functionalization or subfunctionalization (Moriyama and Koshiba-Takeuchi, 2018).

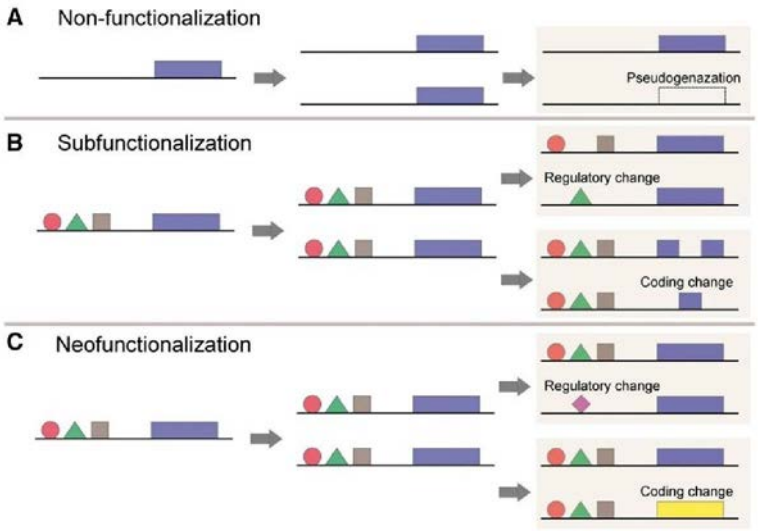


Figure 3. Evolutionary fate of duplicate genes. Rectangles represent genes and circles, triangles and squares represent regulatory regions, such as enhancers (Moriyama and Koshiba-Takeuchi, 2018).

Following a WGD event, there is a brief phase characterized by genome instability, significant gene loss, and heightened levels of nucleotide substitution. During that timeframe, regulatory networks need to be swiftly modified to incorporate the recently replicated genes. Organisms with duplicated proteins exhibit stricter regulation of these duplicates by chemical changes, such as phosphorylation, in comparison to other proteins. This mechanism contributes to the extended viability of duplicated genes, indicating that the preservation of genes is partially programmed (Amoutzias et al., 2010).

2.2.3. Gene fusion and fission

Throughout the process of evolution, genes experience both fusion and fission, resulting in the joining or separation of ORFs (Open Reading Frames). Gene fusions are the result of combining two or more previously separate ORFs. They arise due to chromosomal translocation, chromosomal inversion, or interstitial deletions. Gene fission refers to the reverse process, specifically the division of an open reading frame. Both mechanisms possess the capacity to generate genetic variety and create variations in protein functionalities, also known as neofunctionalization.

Studies have suggested that gene-fission events are infrequent due to the need for numerous simultaneous evolutionary changes at specific sites within an open reading frame. These changes include 1) the gain of a stop codon, 2) the gain of a promoter region, 3) and the appropriation of a start codon.

Gene fusions have been proposed as valuable tools for determining the direction of evolutionary relationships. They are proposed to be useful characters for determining evolutionary relationships since they serve as synapomorphies or cladistic features. Gene fission is driven by at least three mechanisms: separation, degeneration, and duplication. Evidence indicates that gene fission has a significant and previously underestimated impact on gene evolution. Gene fusions, on the other hand, are unstable traits, and their utility in determining evolutionary relationships, without considering gene and species phylogenies, is limited (Leonard and Richards, 2012).

2.2.4. Horizontal gene transfer (HGT)

The advent of high-throughput genome sequencing has greatly enhanced our understanding of genetic changes across the entire spectrum of evolution. Interestingly, not all of this new data conform to a simple vertical-descent model. A significant number of gene trees exhibit incongruences with established phylogenetic relationships of organisms or display discordant topologies. Therefore, numerous genes exhibit patterns across genomes and taxa that cannot be understood by vertical genetic transmission or gene loss. Occasionally, they display distinctive compositional characteristics (such as GC content, codon usage, and nucleotide frequencies) that differ from the rest of the genome, indicating their association with more distantly related species. One possible reason is that genetic material is exchanged between various strains, species, and even kingdoms (Barreiro et al., 2019).

Horizontal gene transfer (HGT) is the process by which genetic material is transferred between unrelated species by multiple mechanisms, including viral infection, transposable elements, and cell fusion (Richards, 2011). Historically, it has been believed that horizontal gene transfer is less common in eukaryotes compared to its frequent occurrence in prokaryotes. While there is ongoing debate over this phenomenon in eukaryotic cells, evidence indicates that horizontal gene transfer plays an active part in the evolution of eukaryotic genomes. HGT serves as an evolutionary shortcut, facilitating the emergence of new capacities in the species involved. Between 0.1% and 2.8% of genes in a normal fungal genome have been influenced by HGT, and up to 5% of clustered biosynthetic genes in fungi have originated through HGT (Barreiro et al., 2019).

There are various potential sources of exogenous DNA for eukaryotic cells such as: (a) lysed organelles derived from endosymbionts (mitochondria and chloroplasts), endosymbionts of prokaryotic or eukaryotic origin, (b) viral infection, (c) bacterial conjugation, and (d) disintegrated phagosomes containing food organisms. The probability of a certain HGT event being successful is determined by multiple factors, that affect the occurrence of HGT in a particular species.

There is increasing evidence that indicates, that viruses can serve as carriers of HGT between prokaryotes and eukaryotes, as well as between different eukaryotic organisms. Eukaryote genomes have been shown to include a plethora of viral genomes and genome fragments, encompassing both retroviruses and other types of viruses (Sibbald et al., 2020).

Transposable elements possess an inherent capacity to relocate within a genome, allowing their transmission to other genomes. Instances of horizontally transferred transposable elements (TEs) have been extensively documented in several organisms, including animals, plants, insects, and fungi. However, large-sale research of TEs in eukaryotes is insufficient, which hinders the ability to draw broad conclusions regarding the magnitude of intergenomic TE mobility and its overall influence on the evolution of eukaryotic genomes (Sibbald et al., 2020).

Laterally acquired genes can give novel functionalities to eukaryotic organisms, influencing their biology and evolutionary path. HGT has been proposed to contribute to the adaptation of eukaryotes to challenging environments, such as intense heat, high levels of salt and anaerobic conditions. HGT also played a role in the development of parasitism and related phenomena, including the transition from a parasitic to a non-parasitic way of life (Sibbald et al., 2020).

The transfer of bacterial DNA to fungi has been frequently documented, with yeasts being the most often reported occurrence. In the example of *S. cerevisiae*, six genes encode enzymes responsible for biotin biosynthesis, with two of them (BIO3 and BIO4) appearing to have been acquired via horizontal gene transfer from a bacterial origin relatively recently. BIO3 and BIO4 correspond to consecutive enzymatic reactions involved in the synthesis of biotin, specifically the production of (7,8-diamino pelargonic acid and dethiobiotin). In other eukaryotes, it appears that a single protein carries out both of these functions. The phylogenetic analysis of these genes indicates that they were obtained separately from distinct prokaryotic donors. Specifically, BIO3

was acquired from gamma-proteobacteria, whereas BIO4 was acquired from alpha-proteobacteria (Barreiro et al., 2019).

In 2018, Li et al. identified seven *R. irregularis* genes that had homologs in multiple plant groups, but not in other fungi, suggesting that they were likely recently transferred from plants to *R. irregularis* (Li et al., 2018).

The transfer of genes between various fungi has also been recorded. A study conducted in 2011 revealed the horizontal gene transfer of a cluster of 23 genes from the *Aspergillus* lineage to *Podospora*. Both of these families are distantly related members of the *Pezizomycotina*, which is a subphylum found within the *Ascomycota*. The 23-gene cluster includes genes that encode the complete sterigmatocystin synthesis pathway, which is responsible for producing the toxic and mutagenic chemical known as sterigmatocystin (Richards, 2011).

2.3. The Pan-Genome Concept

Historically, phylogenetic inference depended on a limited number of commonly used markers, such as ITS, RNA polymerases, translation elongation factors, and mitochondrial genes. The use of genome sequencing has made it a regular practice to reconstruct the evolutionary relationships of species by analyzing significant amount of phylogenetically informative genomic data (McCarthy and Fitzpatrick, 2017). Early in the genome era, scientists have traditionally depended on a singular 'reference' genome per species. This genome served as the fundamental framework for many genetic investigations, such as the examination of variations within and between species. With the decrease in sequencing prices, numerous additional genomes have been sequenced, leading scientists to recognize the insufficiency of relying on a single reference genome for multiple purposes. By collecting a wide range of individuals, it is possible to construct a comprehensive pan-genome (Sherman and Salzberg, 2020).

The first concept of the pan-genome was developed in 2005, during a study on *Streptococcus agalactiae* (Tiwary, 2020). The pan-genome of a species can undergo evolutionary changes due to its lifestyle. Sympatric animals tend to have larger pan-genomes, resulting in a greater degree of variation within the species. On the other hand, ecologically isolated or highly specialized species have smaller pan-genomes. By 2013, the literature has described the pan-genomes of more than 40 species of prokaryotes. In recent years, numerous tools for pan-genome analysis have been developed.

The pan-genome refers to the entire set of genes in the genomes of members of a given species. The term was created to emphasize the fact that individuals or populations of a certain species might exhibit variation in their genetic content (Sibbald et al., 2020). The pan-genome is comprised of three major categories: the core genome, which is found in all individuals of a species; the accessory genome, which is found in only some individuals of a species; and the singleton or unique genome, which is exclusive to a single individual (Figs. 4 and 5). The genes found in the core genome are involved in fundamental cellular metabolic processes, such as housekeeping. Furthermore, the core genome is regarded as a conserved genomic unit used to deduce evolutionary connections between distinct strains. Accessory genes commonly experience gene gain/loss events and are regularly susceptible to horizontal gene transfer in order to promote adaptations in a new ecological niche.

A pan-genome analysis offers three distinct forms of novel information: the size of the core genome, the size of the accessory genome, and the occurrences of gene gain or loss. The effectiveness of a pan-genome study relies on the quality of the reference assembly, the quality of annotation, and the careful selection of suitable individuals for analysis. In Prokaryotes the core part is involved in vertical transmission and homologous recombination, while the variable part is associated with horizontal gene transfer and site-specific recombination (Tiwary, 2020).

A pangenome can be categorized as either open or closed. In an open pangenome, its size continues to grow perpetually as new genomes are added. Therefore, sequencing further strains is likely to result in the discovery of new genes. In contrast, in a close pangenome, the inclusion of other genomes will not result in the identification of novel coding capabilities (Bosi et al., 2015).

The implementation of the pan-genome is influenced by six key factors: 1) the choice of alignment algorithm (such as BLAST or FASTA) and the specific parameters (such as the percentage of identity and sequence length) used to determine the similarity (orthologs, xenologs, or paralogs), 2) the phylogenetic resolution of the target clade (whether it is narrow or wide) 3) the selection or availability of input genomes to represent the target clade, 4) the model used to estimate the number of new genes compared to the number of genomes, 5) The type and quality of sequence annotation (genes, ORFs, or CDSs), 6) the all-against-all level of comparison (sequence similarity

vs phyletic profile of gene presence/absence regardless of sequence similarity (Vernikos et al., 2015).

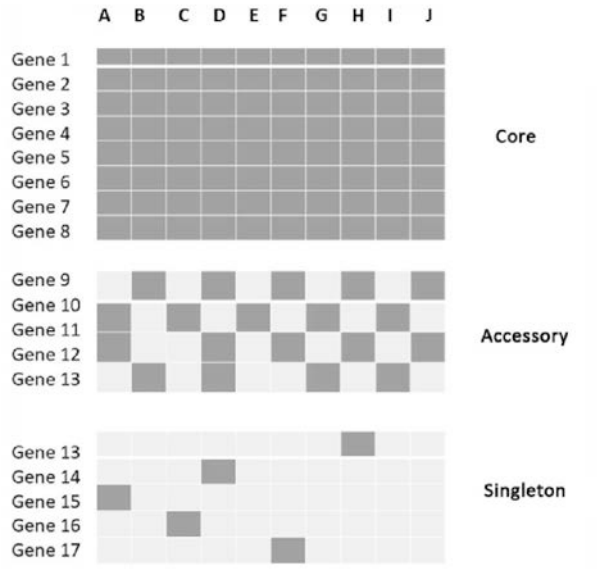


Figure 5. Distribution of individual genes as core genes, accessory genes, and singleton genes in the pan- genome of 10 strains (A–J) of a species (Tiwary, 2020).

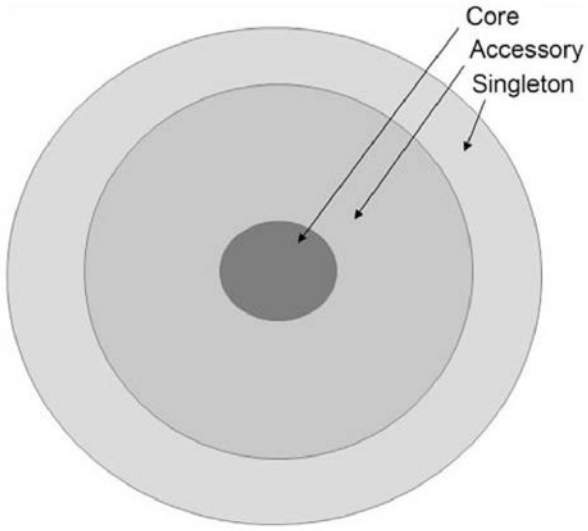


Figure 4. A pan-genome can be classified as the core, accessory, and singleton parts (Tiwary, 2020).

2.3.1. The Pan-Genome Concept in eukaryotes

While the idea of the pan-genome is well accepted in the field of comparative genomics for prokaryotes, it has only recently been used to comparative intraspecific studies for eukaryotes. This is in spite of the consistent detection of variations in the genomic content within a species of eukaryotes, which has been observed from the initial comparative analysis of *Saccharomyces cerevisiae* genomes in the mid-2000s. While horizontal gene transfer (HGT) is considered to be a significant force behind the evolution of gene families and pan-genomes in prokaryotes, HGT happens at much lower frequencies in eukaryotes and is more challenging to identify. Notwithstanding these difficulties, recent studies have examined the variance within different groups of eukaryotes and have provided compelling evidence for the presence of a pan-genome in eukaryotes (McCarthy and Fitzpatrick, 2019). The eukaryote pan-genome evolves differently from prokaryotes because gene duplication is a prevalent mechanism in eukaryotes, while horizontal gene transfer (HGT) is prevalent in prokaryotes. Eukaryotes exhibit genomic differences through single nucleotide polymorphisms (SNPs), copy number variants (CNVs) (i.e., varying number of sequence copies in individuals), and the presence or absence of variants (PAVs) (Tiwary, 2020).

Identifying orthologous genes is essential for constructing the pan-genome of a species, as it is valuable in calculating phylogenetic trees, annotating a genome, and predicting the function of a gene. Orthologous genes are genes that share a common ancestor and have been derived through the process of speciation, while paralogous genes are the result of gene duplication processes. Orthologous genes share a common biological purpose, but paralogous genes have diverse biological activities, even within a specific species. According to the ortholog conjecture, orthologs are expected to have very similar functions because they experience consistent selection pressure, unlike paralogs (Tiwary, 2020).

Pan-genomes have been constructed for several food crop species, such as rice, tomato, soybean and sunflower. Ongoing efforts are being made to conduct pan-genomic analyses in humans and develop population-specific pan-genomes, while also striving to identify as many human SNPs and structural variants as possible (Sherman and Salzberg, 2020).

2.3.2. Pan-genomics of fungi

Fungi have played a major part in the study of eukaryotic genomes. Thanks to efforts like the 1000 Fungal Genomes Project, there has been a significant rise in the amount of genomic data for fungi in the past 5 years. This has allowed for extensive comparative genomics studies of species throughout the fungal kingdom (McCarthy and Fitzpatrick, 2017). Multiple studies on the pan-genome of *Saccharomyces cerevisiae* have revealed compelling evidence of an accessory genome with varying sizes. Additionally, there is significant variation in subterminal regions among multiple strains of *Saccharomyces cerevisiae*, as demonstrated by a comprehensive analysis of genome evolution involving 1011 strains (McCarthy and Fitzpatrick, 2019). Studying the pan-genome of highly variable eukaryotic strains and analyzing their accessory genome provides a more comprehensive knowledge of adaptive evolution. The genomics analysis of *S. cerevisiae* has enhanced our understanding of the evolutionary processes of natural populations in comparison to domesticated strains, as well as during infections and laboratory studies. In addition to *S. cerevisiae*, population genomic studies have also examined the metabolic, genetic, and biogeographic diversity of *Saccharomyces paradoxus*, *Saccharomyces kudriavzevii*, and *Saccharomyces uvarum*. In addition to all other organisms, yeast genome sequences

provide a comprehensive description of their genetic compositions. Comparative genomic studies have significantly enhanced our understanding of the historical and genetic mechanisms involved in the evolution of the studied lineages.

Zymoseptoria tritici, a wheat pathogen responsible for Septoria tritici blotch, has been found to undergo host specialization through gene deletions and chromosomal rearrangements, as revealed by a recent study utilizing pan-genome analysis. In the study mentioned above, the authors utilized five isolates to conduct a pan-genome analysis. They discovered 15,749 nonredundant protein-coding genes as part of the pan-genome, 9,149 protein-coding genes as part of the core genome, and 6,600 protein-coding genes as part of the accessory genome.

The recent advancements in fungal pan-genomics studies have greatly transformed the traditional methods of utilizing biological resources for biotechnology. These studies have been significant in the search and identification of commercially viable biological resources, leading to major scientific and technological progress (Tiwary, 2020).

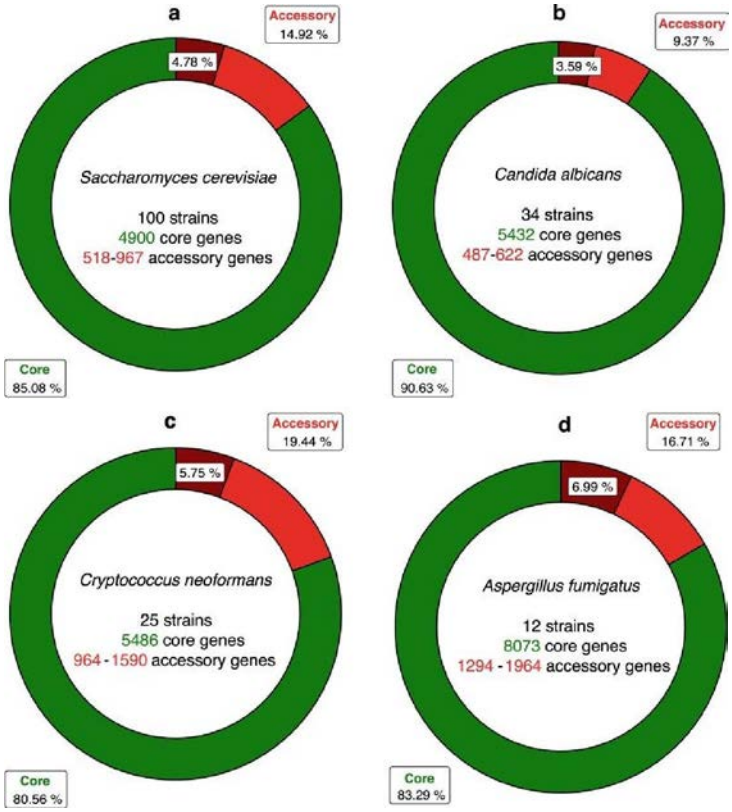


Figure 6. Pan-genomes of four fungal species. The ring charts represent the total number of genes in pan-genomes (McCarthy and Fitzpatrick, 2019).

The goal of this study was to analyze the publicly available *Ascomycetes* genomes via a phylogenomic and pangenome approach in order to better understand the major evolutionary groups that the various species form and the properties of the core and fingerprint proteins of those evolutionary groups. Thus, genes involved in lineage-specific adaptations may be revealed.

3. Materials and Methods

3.1. NCBI Datasets

For the analyses, genomes (Genomic FASTA) and the corresponding proteomes (Protein FASTA) were taken from the NCBI Datasets Database. The NCBI Datasets Database can be queried via two command-line tools (CLI) called "datasets" and "dataformat". Datasets is a tool utilized to query and download biological sequencing data from NCBI databases, encompassing all domains of life, as a package. A genomic data package typically consists of genome, transcript, and protein sequences, together with annotations and one or more data reports. Data packages are downloaded as a zip archive and they include the following files:

- _<assembly_name>_genomic.fna (genomic sequences)
- assembly_data_report.jsonl (data report with genome assembly and annotation metadata)
- dataset_catalog.json (a list of files and file types included in the data package)

For this phylogenomic analysis 91 *Ascomycetes* proteomes (one for each species) and 2 *Basidiomycetes* proteomes (as outgroups) were used (NCBI taxonomy ID: 4890) (Table 1). The corresponding assemblies were retrieved in September 2023. Of these, the first 88 (Table 1) were annotated either "complete", "chromosome" or "scaffold", whereas, for the rest, their assembly level was characterized as "contig".

Table 1. The different species used for the phylogenomic analysis.

	NCBI RefSeq/GenBank assembly	Species
1	GCF_000002495.2	*Pyricularia oryzae*
2	GCF_000002515.2	*Kluyveromyces lactis*
3	GCF_000002525.2	*Yarrowia lipolytica*
4	GCF_000002545.3	*Nakaseomyces glabratus*

5	GCF_000002655.1	*Aspergillus fumigatus*
6	GCF_000002945.1	*Schizosaccharomyces pombe*
7	GCF_000006445.2	*Debaryomyces hansenii*
8	GCF_000011425.1	*Aspergillus nidulans*
9	GCF_000026365.1	*Zygosaccharomyces rouxii*
10	GCF_000026945.1	*Candida dubliniensis*
11	GCF_000027005.1	*Komagataella phaffii*
12	GCF_000091025.4	*Eremothecium gossypii*
13	GCF_000142805.1	*Lachancea thermotolerans*
14	GCF_000143535.2	*Botrytis cinerea*
15	GCF_000146045.2	*Saccharomyces cerevisiae*
16	GCF_000149555.1	*Fusarium verticillioides*
17	GCF_000149955.1	*Fusarium oxysporum f. sp. lycopersici*
18	GCF_000182925.2	*Neurospora crassa*
19	GCF_000182965.3	*Candida albicans*
20	GCF_000184455.2	*Aspergillus oryzae*
21	GCF_000187245.1	*Ogataea parapolymorpha*
22	GCF_000209165.1	*Scheffersomyces stipitis*
23	GCF_000219625.1	*Zymoseptoria tritici*
24	GCF_000226095.1	*Thermothelomyces thermophilus*
25	GCF_000226115.1	*Thermothielavioides terrestris*
26	GCF_000227115.2	*Naumovozyma dairenensis*
27	GCF_000235365.1	*Eremothecium cymbalariae*
28	GCF_000236905.1	*Tetrapisispora phaffii*
29	GCF_000237345.1	*Naumovozyma castellii*
30	GCF_000240135.3	*Fusarium graminearum*
31	GCF_000243375.1	*Torulaspora delbrueckii*
32	GCF_000303195.2	*Fusarium pseudograminearum*
33	GCF_000304475.1	*Kazachstania africana*
34	GCF_000315875.1	*Candida orthopsilosis*
35	GCF_000315915.1	*Tetrapisispora blattae*
36	GCF_000348985.1	*Kazachstania naganishii*
37	GCF_000687475.1	*Ustilaginoidea virens*
38	GCF_001298625.1	*Saccharomyces eubayanus*
39	GCF_001417885.1	*Kluyveromyces marxianus*
40	GCF_001548555.1	*Eremothecium sinecaudum*
41	GCF_001625195.1	*Drechmeria coniospora*
42	GCF_001640025.1	*Sugiyamaella lignohabitans*
43	GCF_001653235.2	*Pochonia chlamydosporia*
44	GCF_001672515.1	*Colletotrichum higginsianum*
45	GCF_001723175.1	*Penicillium oxalicum*

46	GCF_002079055.1	*Saccharomyces paradoxus*
47	GCF_002742065.1	*Cercospora beticola*
48	GCF_003013715.1	*Candida auris*
49	GCF_003054445.1	*Pichia kudriavzevii*
50	GCF_004011695.2	*Ascochyta rabiei*
51	GCF_004337985.1	*Pyricularia pennisetigena*
52	GCF_004355905.1	*Pyricularia grisea*
53	GCF_009556855.1	*Talaromyces marneffei*
54	GCF_011074865.1	*Brettanomyces nanus*
55	GCF_011074885.1	*Brettanomyces bruxellensis*
56	GCF_013085055.1	*Fusarium oxysporum*
57	GCF_013368755.1	*Talaromyces rugulosus*
58	GCF_013402915.1	*Zygotorulaspora mrakii*
59	GCF_014117465.1	*Aspergillus flavus*
60	GCF_014133895.1	*Torulaspora globosa*
61	GCF_016861625.1	*Aspergillus luchuensis*
62	GCF_016861735.1	*Aspergillus chevalieri*
63	GCF_016861865.1	*Aspergillus puulaauensis*
64	GCF_019609905.1	*Fusarium poae*
65	GCF_019915245.1	*Fusarium musae*
66	GCF_020509005.1	*Fulvia fulva*
67	GCF_020623625.1	*Saccharomycodes ludwigii*
68	GCF_022605165.1	*Purpureocillium takamizusanense*
69	GCF_023278565.1	*Colletotrichum lupini*
70	GCF_025433545.1	*Fusarium keratoplasticum*
71	GCF_026873545.1	*Fusarium falciforme*
72	GCF_027921745.1	*Schizosaccharomyces osmophilus*
73	GCF_028009165.1	*Akanthomyces muscarius*
74	GCF_028502605.1	*Trichoderma breve*
75	GCF_900007375.1	*Fusarium venenatum*
76	GCF_900079805.1	*Fusarium fujikuroi*
77	GCF_947241705.1	*Saccharomyces mikatae*
78	GCF_947243775.1	*Saccharomyces kudriavzevii*
79	GCF_000091045.1	*Cryptococcus neoformans*
80	GCF_000328475.2	*Ustilago maydis*
81	GCA_001929475.1	*Neolecta irregularis*
82	GCA_017301755.1	*Pneumocystis wakefieldiae*
83	GCA_017311285.1	*Pneumocystis oryctolagi*
84	GCA_018127085.1	*Pneumocystis sp.'macacae'*
85	GCF_000004155.1	*Schizosaccharomyces cryophilus*
86	GCF_000149845.2	*Schizosaccharomyces japonicus*

87	GCF_000150505.1	*Schizosaccharomyces octosporus*
88	GCF_001661265.1	*Saitoella complicata*
89	GCA_017311265.1	*Pneumocystis canis*
90	GCA_000349005.2	*Pneumocystis murina*
91	GCA_001477535.1	*Pneumocystis jirovecii*
92	GCA_001477545.1	*Pneumocystis carinii*
93	GCA_002105105.1	*Protomyces lactucae-debilis*

3.2. Software

3.2.1. Linux-Ubuntu 20.04 LTS

Ubuntu Linux is currently one of the most popular end-user Linux distributions. As most of Linux software, it is developed as Open Source software. The source code for an application is freely distributed along with the application (Petersen, 2022). Ubuntu exhibits exceptional efficiency and can be adjusted according to specific hardware setups, making it ideal for high-performance computing.

3.2.2. Perl 5

Perl was developed by Larry Wall in 1987 and although it was originally designed for text processing, it has grown into a general-purpose programming language with a rich software development environment. Perl is freely available and freely redistributable (Patwardhan et al., 2002). Perl has become popular with biologists because it is well-suited to several bioinformatics tasks. It is available and runs on all the operating systems found in the average biology lab (Unix and Linux, Macintosh, Windows, VMS, and more). Perl has certain features that simplifies several common bioinformatics tasks. It can deal with information in ASCII text files or flat files, which are the kinds of files in which much important biological data appears, in the GenBank and PDB databases. Perl makes it easy to process and manipulate long sequences such as DNA and proteins (Tisdall, 2001).

3.2.3. Core Proteome Software

To identify the core proteome of the organisms in question, a pipeline that had been developed and applied to bacteria of the genus *Pseudomonas* was used (Nikolaidis

et al., 2020). The pipeline consists of several Python scripts that have been designed to perform pairwise reciprocal DIAMOND search (ultra sensitive mode). The user must specify a reference proteome to conduct the reciprocal DIAMOND against all other proteomes in the dataset. During each comparison, the reference proteome serves as the basis for searching for homologs in the other proteome and this search is repeated in the other direction. Subsequently, proteins that are identified as the best reciprocal matches in each comparison are classified as orthologs.

Next, a matrix is created with the best reciprocal DIAMOND hits. Each row represents a protein from the reference proteome. The columns represent the other proteomes in the selected set and the orthologs of each reference protein if they exist. Following that, a secondary matrix is created, comprising solely the proteins and orthologues that are found in every examined proteome. The second matrix represents the core proteome of the set. By utilizing this particular pipeline, a single reference point is established, allowing for a more rapid execution of the previously mentioned process, compared to other approaches like Orthofinder, which involve conducting all possible comparisons to identify orthologues.

Then, the scripts utilize the MUSCLE software to generate multiple alignments for the recognized core orthologues. These alignments are then combined into a super-alignment and any poorly matched sections are removed using the Gblocks software. The final alignment is utilized for the generation of a Maximum Likelihood phylogenetic tree, using the software IQ tree2.

3.2.4. Interactive Tree of Life (iTOL)

The Interactive Tree Of Life (https://itol.embl.de) is an online tool designed for the display, editing, and annotation of phylogenetic and other types of trees. It is free and accessible to all users. iTOL is compatible with widely used phylogenetic tree formats, including Newick, Nexus, and phyloXML. iTOL offers a range of features that are often seen in other phylogenetic tree viewers and different tree display forms, including phylograms or cladograms, rooted or unrooted, rectangular or radial. Moreover, iTOL v5 provides a range of annotation options and enhanced functionality for visualizing different types of datasets. The majority of these datasets can be uploaded as plain text files. Finally, one of the main purposes of iTOL is to generate

high-quality graphics for publishing or for inclusion into other papers (Letunic and Bork, 2021).

3.2.5. eggNOG-mapper v2

EggNOG-mapper is a program designed for functional annotation using precomputed orthology assignments, that can handle large-scale genomic and metagenomic datasets. The inference of gene function through orthology, as opposed to homology detection, is often regarded as the most dependable method for transferring functional information between molecular sequences. This is because orthologs are expected to preserve function more frequently than paralogs. EggNOG-mapper utilizes the eggNOG database, which contains orthologous groups (OGs) encompassing a wide range of bacterial, archaeal, and eukaryotic organisms. To achieve this, it utilizes the precalculated phylogenies determined for each orthologous group to effectively improve orthology assignments and hence reduce the transfer of annotations from potential in-paralogs. This method was shown to produce more precise predictions compared to homology-based approaches, while maintaining computing efficiency at the genomic and metagenomic level. EggNOG-mapper offers a range of annotation sources to generate an in-depth report of functional annotations for each query. These sources include predicted protein name, KEGG pathways, modules, and orthologs, Gene Ontology labels, EC numbers, BiGG reactions, CAZy terms, COG functional categories, eggNOG OGs and free text descriptions at all taxonomic levels (Cantalapiedra et al., 2021).

The Clusters of Orthologous Groups of proteins (COGs) database attempts to phylogenetically classify the proteins encoded in whole genomes. Every COGs consists of proteins that are inferred to be orthologs. The current database has 4,877 COGs, which include proteins from 1,309 organisms.

4. Results – Discussion

4.1. Phylogenomic analysis

For this phylogenomic analysis, 91 *Ascomycetes* genomes (Genomic FASTA) and proteomes (Protein FATSTA) were used (NCBI taxonomy ID: 4890). The

proteomes were retrieved from the NCBI Assembly database in September 2023. Of these, the first 88 (Table 1) were annotated either "complete", "chromosome" or "scaffold", whereas, for the rest, their assembly level was characterized as "contig".

The 91 *Ascomycetes* proteomes were entered into the pipeline (Nikolaidis et al, 2020), with the *Saccharomyces cerevisiae* proteome as a reference (GCF_000146045.2), in order to identify the orthologous proteins. The analysis of the genus revealed that its proteome consists of 489 core proteins.

The orthologous groups were then used to generate multiple alignments that were then combined into a super-alignment. The final alignment, which comprised of 101,679 amino acid sites, was utilized for the generation of a Maximum Likelihood phylogenetic tree, using the software IQ tree2 (Fig. 7). Selecting an outgroup is essential, because it enables the determination of the direction of character change on a rooted phylogeny, facilitating the understanding of trait evolution along the phylogeny. The outgroup is a closely related, but distinct group that is used as a reference point for comparing and establishing the root of the phylogeny of the ingroup. Phylogeneticists typically include additional outgroups in their phylogenetic analysis. It is beneficial to include multiple outgroups, since they provide a robust phylogeny, which helps to counter weak outgroup candidates and examine the proposed monophyly of the ingroup (Dissanayake, 2020). Therefore, the selection of the outgroups was based on the current understanding of the *Opisthokonta* phylogeny and the following criteria: the outgroups must belong to the clade *Basidiomycetes* (sister clade to *Ascomycetes*), they must be unicellular and they must have a genome of relatively similar size to that of the *Ascomycetes*. Of the genomes available at NCBI, two met the criteria, *Cryptococcus neoformans* (*Agaricomycotina*) and *Ustilago maydis* (*Ustilaginomycotina*).

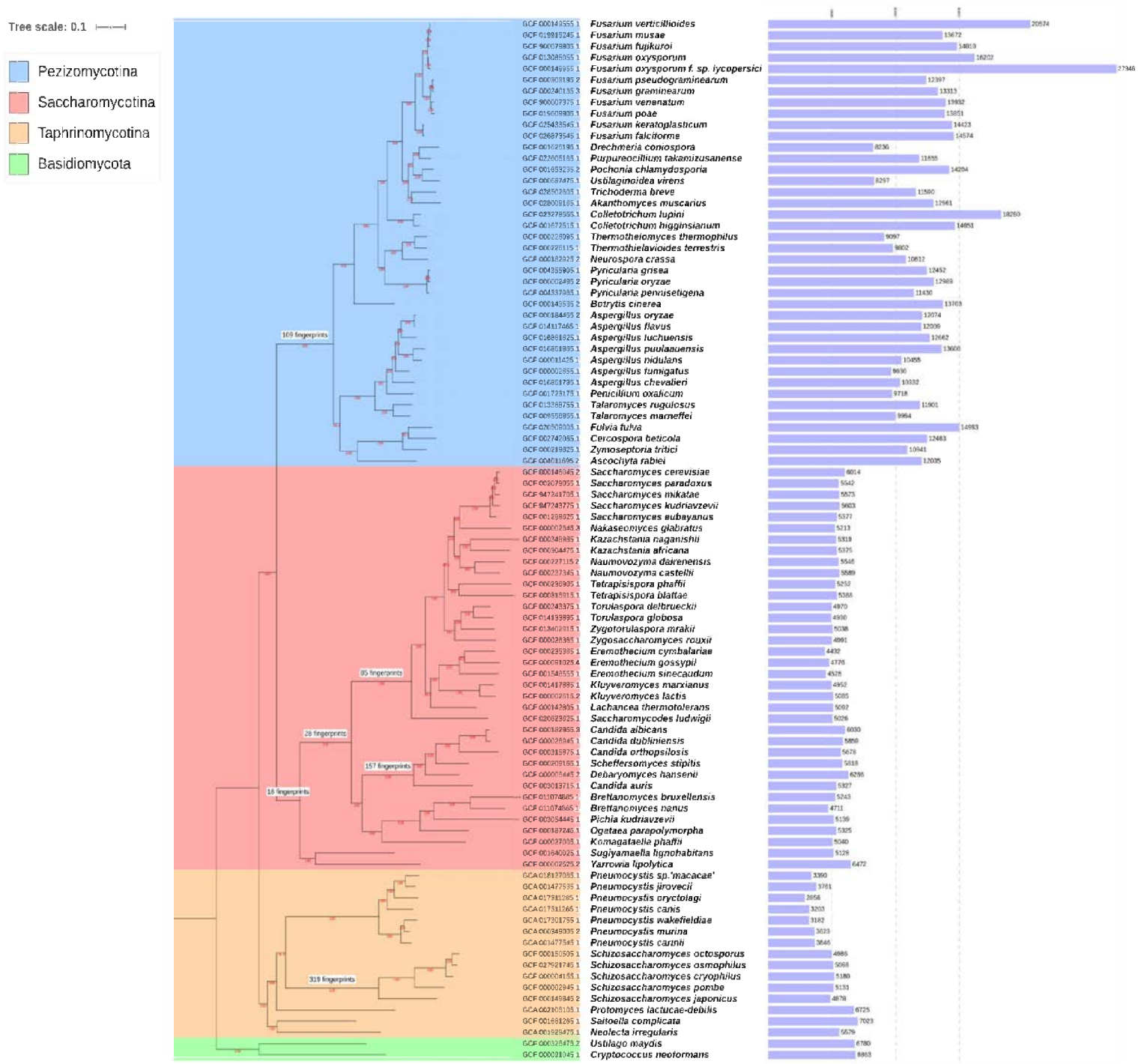


Figure 7. Phylogenomic Maximum Likelihood tree of the 91 Ascomycetes proteomes. Next to the NCBI code are depicted: 1) the species name, 2) the number of proteins per species. Bootstrap values are noted by number below the respective branches.

4.2. Core proteins of the *Ascomycetes* genus

We then calculated, using the hypergeometric test, the statistical significance of the fold enrichment and depletion of the various functional categories in the core proteins of the *Ascomycetes* genus. The results are summarized in Table 3. The

assignment of functional category of the proteins was based on the eggNOG-mapper database, which contains orthologs groups (OGs) encompassing a wide range of bacterial, archaeal, and eukaryotic organisms. EggNOG-mapper annotated each query to COG functional categories (Table 2).

Table 2. COG categories and their functional annotation.

Category Symbol	Annotation
A	RNA processing and modification
B	Chromatin structure and dynamics
C	Energy production and conversion
D	Cell cycle control, cell division, chromosome partitioning
E	Amino acid transport and metabolism
F	Nucleotide transport and metabolism
G	Carbohydrate transport and metabolism
H	Coenzyme transport and metabolism
I	Lipid transport and metabolism
J	Translation, ribosomal structure and biogenesis
K	Transcription
L	Replication, recombination and repair
M	Cell wall/membrane/envelope biogenesis
N	Cell motility
O	Posttranslational modification, protein turnover, chaperones
P	Inorganic ion transport and metabolism
Q	Secondary metabolites biosynthesis, transport and catabolism
R	General function prediction only
S	Function unknown
T	Signal transduction mechanisms
U	Intracellular trafficking, secretion, and vesicular transport
V	Defense mechanisms
W	Extracellular structures
X	Mobilome: prophages, transposons
Y	Nuclear structure
Z	Cytoskeleton

Table 3. Results of the hypergeometric test. To the left of the forward slash is the p-value and to the right is the fold change of the subset relative to the whole reference proteome. Categories with fold change values that are statistically significant (p-value < 0.05) and are enriched (fold change > 1) are marked with green, while categories with fold change values that are statistically significant (p-value < 0.05) and are depleted (fold change < 1) are marked with red.

COG Category	Core Proteins
A	6.76e-09/2.168

B	1.35e-04/2.513
C	0.264/0.84
D	1.61e-02/1.662
E	7.12e-04/0.337
F	4.40e-02/1.64
G	0.065/0.624
H	4.21e-03/1.816
I	2.19e-02/1.634
J	1.30e-06/1.796
K	0.155/0.841
L	0.093/1.27
M	1.17e-02/2.523
N	0.156/6.149
O	5.65e-04/1.548
P	0.233/0.777
Q	0.258/0.691
S	2.05e-30/0.267
T	0.241/0.832
U	0.084/1.19
V	0.408/0.512
W	0.346/2.46
Y	3.30e-03/3.075
Z	0.102/1.447

The functional categories of RNA processing and modification (A), chromatin structure and dynamics (B), cell cycle control, cell division, chromosome partitioning (D), nucleotide transport and metabolism (F), coenzyme transport and metabolism (H), lipid transport and metabolism (I), translation, ribosomal structure and biogenesis (J), cell wall/membrane/envelope biogenesis (M), posttranslational modification, protein turnover, chaperones (O), and nuclear structure (Y) are enriched in the core proteome of *Ascomycetes*, as expected, given that they are proteins involved in essential cellular processes. The category of unknown function (S) is significantly depleted. This is also expected, as the study around the core proteome of the genus *Ascomycetes* is extensive, so many of the orthologs have been well studied by researchers and are characterized.

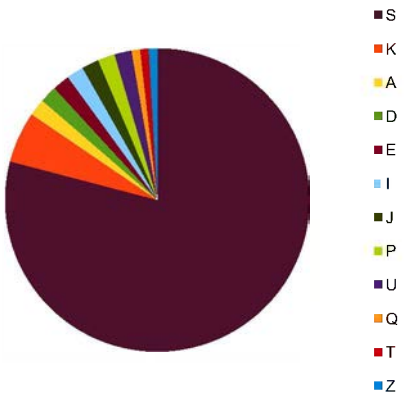
4.3. Core and fingerprint proteins of the *Pezizomycotina* clade

In order to identify the core proteomes of the major evolutionary subgroups, we used a reference proteome from within each subgroup. At the same time, the total number of proteins that appeared in all members of one subgroup but in none of the other subgroups was calculated, giving the fingerprint proteins.

For *Pezizomycotina*, *Fusarium oxysporum* was used as a reference proteome, 40 genomes were analyzed and 2476 core proteins were found. Additionally, 109 fingerprint proteins were found and Table 4 below shows their assignment to functional categories. The full fingerprint proteins list is available at Appendix Table 14.

Table 4. Pezizomycotina fingerprint proteins.

COG Category	Count	Count%
S	87	79.82
K	6	5.5
A	2	1.83
D	2	1.83
E	2	1.83
I	2	1.83
J	2	1.83
P	2	1.83
U	2	1.83
Q	1	0.92
T	1	0.92
Z	1	0.92



We then calculated, using the hypergeometric test, the statistical significance of the fold enrichment and depletion of the functional categories in these fingerprint proteins. The results are summarized in Table 5.

Table 5. Results of the hypergeometric test. To the left of the forward slash is the p-value and to the right is the fold change of the subset relative to the whole reference proteome. When the fold change is equal to 0, it means that no such protein was found. Categories with fold change values that are statistically significant (p-value < 0.05) and are enriched (fold change > 1) are marked with green, while categories with fold change values that are statistically significant (p-value < 0.05) and are depleted (fold change < 1) are marked with red.

COG Category	*Pezizomycotina* Fingerprints

A	0.657/0.965
B	0.275/0.0
C	1.58e-02/0.0
D	0.295/1.846
E	0.108/0.391
F	0.347/0.0
G	1.19e-03/0.0
H	0.16/0.0
I	0.404/0.652
J	0.468/0.716
K	0.23/1.457
L	0.111/0.0
M	0.273/0.0
N	0.96/0.0
O	5.53e-03/0.0
P	0.444/0.691
Q	2.08e-02/0.177
S	1.60e-10/1.587
T	0.155/0.305
U	0.312/0.568
V	0.501/0.0
W	0.947/0.0
Y	0.828/0.0
Z	0.735/0.998

The category of unknown function (S) is significantly enriched, while categories such as energy production and conversion (C), carbohydrate transport and metabolism (G) and posttranslational modification, protein turnover, chaperones (O) are significantly depleted.

4.5. Core and fingerprint proteins of the *Taphrinomycotina* clade

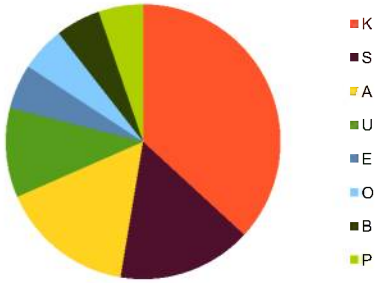
For *Taphrinomycotina*, *Schizosaccharomyces pombe* was used as a reference proteome, 14 genomes were analyzed and 1468 core proteins and 0 fingerprint proteins were found.

4.6. Core and fingerprint proteins of the *Saccharomycotina* clade

For *Saccharomycotina*, *Saccharomyces cerevisiae* was used as a reference proteome, 36 genomes were analyzed and 1692 core proteins were found. Additionally, 18 fingerprint proteins were found and Table 6 below shows their assignment to functional categories. The full fingerprint proteins list is available at Appendix Table 15.

Table 6. Saccharomycotina fingerprint proteins

COG Category	Count	Count%
K	7	38.89
S	3	16.67
A	3	16.67
U	2	11.11
E	1	5.56
O	1	5.56
B	1	5.56
P	1	5.56



For *Saccharomycotina* fingerprint proteins, the category of unknown function (S) represents only the 16.7% of them, as might be expected, given that *Saccharomyces cerevisiae* is one of the most well studied organisms. As mentioned before, it is anticipated that lineage-specific fingerprints may be involved in adaptations of that lineage. Our analyses showed fingerprint proteins that are involved in oxidative stress response, such as the Pet123 protein, which has been shown to functionally interact with Pet122, a translational activator for cytochrome c oxidase subunit III (Desai et al., 2017). Additionally, various types of organisms use sporulation as a way of adapting to changes in their individual environments, as it facilitates survival under adverse growing conditions and dispersal to new favorable habitats (Huang and Hull, 2017). Our analyses revealed a fingerprint protein (Ssp2p) that localizes to the spore wall and is required for sporulation at a point after meiosis II and during spore wall formation. Cells are also able to swiftly respond and adapt to variations in nutrients due to the fast and dynamic turnover of nutrient transporters caused by endocytosis (Barata-Antunes et al., 2021). Our analyses showed a fingerprint protein that is a putative divalent metal

ion transporter (SMF3) as well as a fingerprint protein that is required for normal actin organization and endocytosis (Scd5p).

We then calculated, using the hypergeometric test, the statistical significance of the fold enrichment and depletion of the functional categories in these fingerprint proteins. The results are summarized in Table 7.

Table 7. Results of the hypergeometric test. To the left of the forward slash is the p-value and to the right is the fold change of the subset relative to the whole reference proteome. When the fold change is equal to 0, it means that no such protein was found. Categories with fold change values that are statistically significant (p-value < 0.05) and are enriched (fold change > 1) are marked with green, while categories with fold change values that are statistically significant (p-value < 0.05) and are depleted (fold change < 1) are marked with red.

COG Category	*Saccharomycotina* Fingerprints
A	0.072/3.047
B	0.245/3.593
C	0.467/0.0
D	0.638/0.0
E	0.488/1.526
F	0.728/0.0
G	0.549/0.0
H	0.636/0.0
I	0.648/0.0
J	0.25/0.0
K	3.59e-04/4.706
L	0.385/0.0
M	0.889/0.0
N	0.994/0.0
O	0.62/0.765
P	0.411/1.92
Q	0.764/0.0
S	0.248/0.622
T	0.442/0.0
U	0.558/1.095
V	0.93/0.0
W	0.985/0.0
Y	0.908/0.0
Z	0.697/0.0

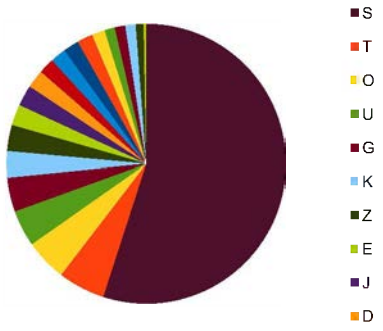
The category of transcription (K) appears to be the only significantly enriched category for the *Saccharomycotina* clade.

4.7. Core and fingerprint proteins of the *Schizosaccharomyces* clade

For *Schizosaccharomyces* (node I117, Fig.7), *Schizosaccharomyces pombe* was used as a reference proteome and 5 genomes were analyzed, where 4103 core proteins were found. Additionally, 319 fingerprint proteins were found and the Table 8 shows their assignment to functional categories. The full fingerprint proteins list is available at Appendix Table 16.

Table 8. Schizosaccharomyces fingerprint proteins

COG Category	Count	Count%
S	182	57.05
T	18	5.64
O	16	5.02
U	14	4.39
G	13	4.08
K	10	3.13
Z	10	3.13
E	8	2.51
J	8	2.51
D	7	2.19
C	6	1.88
H	6	1.88
I	6	1.88
L	6	1.88
B	5	1.57
A	4	1.25
M	4	1.25
P	4	1.25
Q	3	0.94
F	1	0.31



We then calculated, using the hypergeometric test, the statistical significance of the fold enrichment and depletion of the functional categories in these fingerprint proteins. The results are summarized in Table 9.

Table 9. Results of the hypergeometric test. To the left of the forward slash is the p-value and to the right is the fold change of the subset relative to the whole reference proteome. When the fold change is equal to 0, it means that no such protein was found. Categories with fold change values that are statistically

significant (p-value < 0.05) and are enriched (fold change > 1) are marked with green, while categories with fold change values that are statistically significant (p-value < 0.05) and are depleted (fold change < 1) are marked with red.

COG Category	*Schizosaccharomyces* Fingerprints
A	9.06e-05/0.234
B	0.337/0.752
C	0.071/0.548
D	0.4/0.84
E	0.1/0.628
F	3.81e-02/0.204
G	0.364/1.13
H	0.152/0.631
I	0.332/0.772
J	5.60e-06/0.295
K	2.43e-03/0.458
L	2.10e-03/0.367
M	0.249/1.569
N	0.88/0.0
O	4.46e-02/0.667
P	0.073/0.48
Q	0.24/0.588
S	5.17e-32/2.126
T	0.268/1.177
U	1.95e-03/0.508
V	0.335/0.0
W	0.825/0.0
Y	0.243/0.0
Z	0.355/1.166

The category of unknown function (S) appears to be significantly enriched, while categories such as RNA processing and modification (A), nucleotide transport and metabolism (F), translation, ribosomal structure and biogenesis (J) and posttranslational modification, protein turnover, chaperones (O) are depleted.

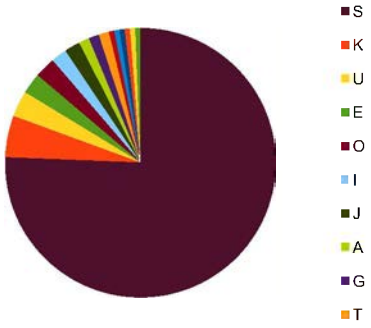
4.8. Core and fingerprint proteins of the C*andida* clade (node I13)

For the *Candida* clade (node I13, Fig.7), C*andida albicans* was used as a reference proteome with 6 genomes were analyzed and 4038 core proteins were found. Additionally, 157 fingerprint proteins were found and Table 10 blow shows their

assignment to functional categories. The full fingerprint proteins list is available at Appendix Table 17.

Table 10. Candida fingerprint proteins

COG Category	Count	Count%
S	121	77.07
K	8	4.46
U	5	3.18
E	4	2.55
O	4	2.55
I	3	1.91
J	3	1.91
A	2	1.27
G	2	1.27
T	2	1.27
C	1	0.64
F	1	0.64
M	1	0.64
P	1	0.64
Y	1	0.64
L	1	0.64



We then calculated, using the hypergeometric test, the statistical significance of the fold enrichment and depletion of the functional categories in these fingerprint proteins. The results are summarized in Table 11.

Table 11. Results of the hypergeometric test. To the left of the forward slash is the p-value and to the right is the fold change of the subset relative to the whole reference proteome. When the fold change is equal to 0, it means that no such protein was found. Categories with fold change values that are statistically significant (p-value < 0.05) and are enriched (fold change > 1) are marked with green, while categories with fold change values that are statistically significant (p-value < 0.05) and are depleted (fold change < 1) are marked with red.

COG Category	*Candida* Fingerprints
A	2.27e-02/0.279
B	0.08/0.0
C	1.45e-02/0.166
D	2.03e-02/0.0
E	0.149/0.559
F	0.407/0.505
G	0.055/0.331
H	4.43e-02/0.0

I	0.25/0.594
J	1.30e-02/0.319
K	0.383/0.842
L	2.72e-02/0.187
M	0.662/0.835
N	0.974/0.0
O	1.29e-02/0.368
P	0.067/0.233
Q	2.26e-02/0.0
S	1.17e-26/2.154
T	2.37e-02/0.281
U	2.99e-02/0.452
V	0.464/0.0
W	0.876/0.0
Y	0.571/1.2
Z	0.056/0.0

The category of unknown function (S) is significantly enriched, while many categories such as energy production and conversion (C), cell cycle control, cell division, chromosome partitioning (D), coenzyme transport and metabolism (H) and secondary metabolites biosynthesis, transport and catabolism (Q) are significantly depleted.

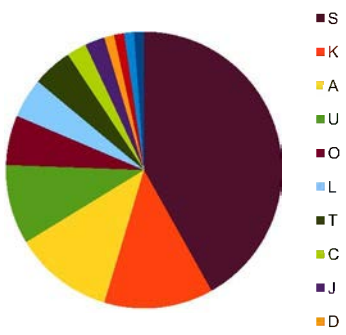
4.8. Core and fingerprint proteins of the *Saccharomyces* clade (node I10)

For the *Saccharomyces* clade (node I10, Fig.7), *Saccharomyces cerevisiae* was used as a reference proteome with 23 genomes analyzed and 2861 core proteins were found. Additionally, 85 fingerprint proteins were found and Table 12 shows their assignment to functional categories. The full fingerprint proteins list is available at Appendix Table 18.

Table 12. Saccharomyces fingerprint proteins

COG Category	Count	Count%
S	36	42.35
K	11	12.94
A	10	11.76
U	8	9.41

O	5	5.88
L	4	4.71
T	4	4.71
C	2	2.35
J	2	2.35
D	1	1.18
G	1	1.18
P	1	1.18
Z	1	1.18



We then calculated, using the hypergeometric test, the statistical significance of the fold enrichment and depletion of the functional categories in these fingerprint proteins. The results are summarized in Table 13.

Table 13. Results of the hypergeometric test. To the left of the forward slash is the p-value and to the right is the fold change of the subset relative to the whole reference proteome. When the fold change is equal to 0, it means that no such protein was found. Categories with fold change values that are statistically significant (p-value < 0.05) and are enriched (fold change > 1) are marked with green, while categories with fold change values that are statistically significant (p-value < 0.05) and are depleted (fold change < 1) are marked with red.

COG Category	*Saccharomyces* Fingerprints
A	1.69e-02/2.151
B	0.263/0.0
C	0.309/0.568
D	0.376/0.478
E	4.18e-02/0.0
F	0.221/0.0
G	0.227/0.359
H	0.117/0.0
I	0.127/0.0
J	4.34e-02/0.318
K	0.09/1.566
L	0.552/0.913
M	0.573/0.0
N	0.972/0.0
O	0.41/0.81
P	0.289/0.407
Q	0.279/0.0
S	1.29e-03/1.58
T	0.523/1.064
U	0.501/0.928
V	0.71/0.0
W	0.931/0.0

Y	0.633/0.0
Z	0.495/0.595

The categories of unknown function (S) and RNA processing and modification (A) appear to be enriched, while the categories of amino acid transport and metabolism translation (E) and ribosomal structure and biogenesis (J) are depleted.

Below is a table summarizing the fingerprints proteins from each evolutionary group in each COG category:

FINGERPRINTS PROTEINS																									
Phylogenomic Group	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
Candida (Node I13)	2	0	1	0	4	1	2	0	3	3	8	1	1	0	4	1	0	0	121	2	5	0	0	0	1	0
Saccharomyces (Node I10)	10	0	2	1	0	0	1	0	0	2	11	4	0	0	5	1	0	0	36	4	8	0	0	0	0	1
Pezizomycotina	2	0	0	2	2	0	0	0	2	2	6	0	0	0	0	2	1	0	87	1	2	0	0	0	0	1
Saccharomycotina	3	1	0	0	1	0	0	0	0	0	7	0	0	0	1	1	0	0	3	0	2	0	0	0	0	0
Schizosaccharomyces	4	5	6	7	8	1	13	6	6	8	10	6	4	0	16	4	3	0	182	18	14	0	0	0	0	10
Taphrinomycotina	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

All of the analyses in our study demonstrate, that the vast majority of the fingerprints proteins still remain of unknown function (Category S). A remarkable observation was that although *Pezizomycotina* is a large clade, with 40 different species (Fig. 7) we identified 109 fingerprint proteins. This high number of lineage-specific fingerprints and the long branch in the phylogenomic tree strongly indicate that this lineage has significantly diverged from the other *Ascomycetes*. For *Schizosaccharomyces* 319 fingerprint proteins were found, but only for organisms of the same genus (5 different species). Finally, the total number of the core proteins of the evolutionary subgroups within *Ascomycota* was decreasing as the number of genomes tested was increasing, as expected.

5. Conclusions

This work involved the analysis of genetic data obtained from 91 *Ascomycetes*, along with 2 taxa that were used as outgroups. We deduced a thorough phylogeny based on the orthologous groups and determined the core proteins of the group and of the

various phylogenetic sub-groups. For these sub-groups we also identified proteins that distinguish them (fingerprints) form the others. As mentioned before, the study of fingerprint proteins is important, because it can reveal proteins that may be involved in lineage-specific adaptations and thus offer an important evolutionary advantage, that gives the members of a particular group the ability to thrive in their respective environments. The majority of the fingerprint proteins found in this study remain of unknown function, indicating that there is a great need for biochemical and functional characterization of these proteins in order to better understand the adaptations of the various phylogenetic groups.

6. Bibliography

Amoutzias, G.D., He, Y., Gordon, J., Mossialos, D., Oliver, S.G., Van de Peer, Y., 2010. Posttranslational regulation impacts the fate of duplicated genes. Proc Natl Acad Sci U S A 107, 2967–2971. https://doi.org/10.1073/pnas.0911603107

Barata-Antunes C, Alves R, Talaia G, Casal M, Gerós H, Mans R, Paiva S. Endocytosis of nutrient transporters in fungi: The ART of connecting signaling and trafficking. Comput Struct Biotechnol J. 2021 Mar 19;19:1713-1737. doi: 10.1016/j.csbj.2021.03.013. PMID: 33897977; PMCID: PMC8050425.

Barone, G.D., Emmerstorfer-Augustin, A., Biundo, A., Pisano, I., Coccetti, P., Mapelli, V., Camattari, A., 2023. Industrial Production of Proteins with Pichia pastoris—Komagataella phaffii. Biomolecules 13, 441. https://doi.org/10.3390/biom13030441

Barreiro, C., Gutierrez, S., Olivera, E., 2019. Fungal Horizontal Gene Transfer: A History Beyond the Phylogenetic Kingdoms. pp. 315–336. https://doi.org/10.1007/978-3-030-21862-1_13

Batt, C. A., 2014. Pichia pastoris, in: Batt, Carl A., Tortorello, M.L. (Eds.), Encyclopedia of Food Microbiology (Second Edition). Academic Press, Oxford, pp. 42–46. https://doi.org/10.1016/B978-0-12-384730-0.00254-8

Bosi, E., Fani, R., Fondi, M., 2015. Defining Orthologs and Pangenome Size Metrics, in: Mengoni, A., Galardini, M., Fondi, M. (Eds.), Bacterial Pangenomics: Methods and Protocols, Methods in Molecular Biology. Springer, New York, NY, pp. 191–202. https://doi.org/10.1007/978-1-4939-1720-4_13

Buckley, M., 2008. The Fungal Kingdom: diverse and essential roles in earth's ecosystem: This report is based on a colloquium, sponsored by the American Academy of Microbiology, convened November 2–4, 2007 in Tucson, Arizona, American Academy of Microbiology Colloquia Reports. American Society for Microbiology, Washington (DC).

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Molecular Biology and Evolution 38, 5825–5829. https://doi.org/10.1093/molbev/msab293

Copley, S.D., 2020. Evolution of new enzymes by gene duplication and divergence. The FEBS Journal 287, 1262–1283. https://doi.org/10.1111/febs.15299

Desai N, Brown A, Amunts A, Ramakrishnan V. The structure of the yeast mitochondrial ribosome. Science. 2017 Feb 3;355(6324):528-531. doi: 10.1126/science.aal2415. PMID: 28154081; PMCID: PMC5295643.

Dissanayake, A., 2020. Applied aspects of methods to infer phylogenetic relationships amongst fungi. Mycosphere 11, 2652–2676. https://doi.org/10.5943/mycosphere/11/1/18

Gautam, A.K., Verma, R.K., Avasthi, S., Sushma, Bohra, Y., Devadatha, B., Niranjan, M., Suwannarach, N., 2022. Current Insight into Traditional and Modern Methods in Fungal Diversity Estimates. Journal of Fungi 8, 226. https://doi.org/10.3390/jof8030226

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G., 1996. Life with 6000 genes. Science 274, 546, 563–567. https://doi.org/10.1126/science.274.5287.546

Huang M, Hull CM. Sporulation: how to survive on planet Earth (and beyond). Curr Genet. 2017 Oct;63(5):831-838. doi: 10.1007/s00294-017-0694-7. Epub 2017 Apr 18. PMID: 28421279; PMCID: PMC5647196.

James, T.Y., Stajich, J.E., Hittinger, C.T., Rokas, A., 2020. Toward a Fully Resolved Fungal Tree of Life. Annual Review of Microbiology 74, 291–313. https://doi.org/10.1146/annurev-micro-022020-051835

Karathia, H., Vilaprinyo, E., Sorribas, A., Alves, R., 2011. Saccharomyces cerevisiae as a Model Organism: A Comparative Study. PLOS ONE 6, e16015. https://doi.org/10.1371/journal.pone.0016015

Kellis, M., Birren, B.W., Lander, E.S., 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature 428, 617–624. https://doi.org/10.1038/nature02424

Kim, J., Sudbery, P., 2011. Candida albicans, a major human fungal pathogen. J Microbiol. 49, 171–177. https://doi.org/10.1007/s12275-011-1064-7

Lallemand, T., Leduc, M., Landès, C., Rizzon, C., Lerat, E., 2020. An Overview of Duplicated Gene Detection Methods: Why the Duplication Mechanism Has to Be Accounted for in Their Choice. Genes 11, 1046. https://doi.org/10.3390/genes11091046

Leonard, G., Richards, T.A., 2012. Genome-scale comparative analysis of gene fusions, gene fissions, and the fungal tree of life. Proc Natl Acad Sci U S A 109, 21402–21407. https://doi.org/10.1073/pnas.1210909110

Letunic, I., Bork, P., 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research 49, W293–W296. https://doi.org/10.1093/nar/gkab301

Li, M., Zhao, J., Tang, N., Sun, H., Huang, J., 2018. Horizontal Gene Transfer From Bacteria and Plants to the Arbuscular Mycorrhizal Fungus Rhizophagus irregularis. Frontiers in Plant Science 9.

Li, Y., Steenwyk, J.L., Chang, Y., Wang, Y., James, T.Y., Stajich, J.E., Spatafora, J.W., Groenewald, M., Dunn, C.W., Hittinger, C.T., Shen, X.-X., Rokas, A., 2021. A genome-scale phylogeny of the kingdom Fungi. Curr Biol 31, 1653-1665.e5. https://doi.org/10.1016/j.cub.2021.01.074

Lin, P., Kook, M., Yi, T.-H., Yan, Z.-F., 2023. Current Fungal Taxonomy and Developments in the Identification System. Curr Microbiol 80, 375. https://doi.org/10.1007/s00284-023-03514-7

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., Ravikesavan, R., 2013. Gene duplication as a major force in evolution. J Genet 92, 155–161. https://doi.org/10.1007/s12041-013-0212-8

Makino, T., Hokamp, K., McLysaght, A., 2009. The complex relationship of gene duplication and essentiality. Trends in Genetics 25, 152–155. https://doi.org/10.1016/j.tig.2009.03.001

Mayer, F.L., Wilson, D., Hube, B., 2013. Candida albicans pathogenicity mechanisms. Virulence 4, 119–128. https://doi.org/10.4161/viru.22913

McCarthy, C.G.P., Fitzpatrick, D.A., 2019. Pan-genome analyses of model fungal species. Microb Genom 5, e000243. https://doi.org/10.1099/mgen.0.000243

McCarthy, C.G.P., Fitzpatrick, D.A., 2017. Multiple Approaches to Phylogenomic Reconstruction of the Fungal Kingdom. Adv Genet 100, 211–266. https://doi.org/10.1016/bs.adgen.2017.09.006

Moriyama, Y., Koshiba-Takeuchi, K., 2018. Significance of whole-genome duplications on the emergence of evolutionary novelties. Briefings in Functional Genomics 17, 329–338. https://doi.org/10.1093/bfgp/ely007

Mukaremera, L., Lee, K.K., Mora-Montes, H.M., Gow, N.A.R., 2017. Candida albicans Yeast, Pseudohyphal, and Hyphal Morphogenesis Differentially Affects Immune Recognition. Frontiers in Immunology 8.

Naranjo-Ortiz, M.A., Gabaldón, T., 2019. Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. Biological Reviews 94, 2101–2137. https://doi.org/10.1111/brv.12550

Nikolaidis, M., Mossialos, D., Oliver, S.G., Amoutzias, G.D., 2020. Comparative Analysis of the Core Proteomes among the Pseudomonas Major Evolutionary Groups Reveals Species-Specific Adaptations for Pseudomonas aeruginosa and Pseudomonas chlororaphis. Diversity 12, 289. https://doi.org/10.3390/d12080289

Oakley, B.R., 2013. Aspergillus nidulans, in: Maloy, S., Hughes, K. (Eds.), Brenner's Encyclopedia of Genetics (Second Edition). Academic Press, San Diego, pp. 212–215. https://doi.org/10.1016/B978-0-12-374984-0.00104-2

Parapouli, M., Vasileiadis, A., Afendra, A.-S., Hatziloukas, E., 2020. Saccharomyces cerevisiae and its industrial applications. AIMS Microbiol 6, 1–31. https://doi.org/10.3934/microbiol.2020001

Patwardhan, N., Siever, E., Spainhour, S., 2002. Perl. O'Reilly Media, Inc.

Pérez, J.C., 2021. Fungi of the human gut microbiota: Roles and significance. International Journal of Medical Microbiology 311, 151490. https://doi.org/10.1016/j.ijmm.2021.151490

Petersen, R., 2022. Ubuntu 22.04 LTS Desktop: Applications and Administration. Surfing Turtle Press.

Richards, T.A., 2011. Genome Evolution: Horizontal Movements in the Fungi. Current Biology 21, R166–R168. https://doi.org/10.1016/j.cub.2011.01.028

Richards, T.A., Leonard, G., Wideman, J.G., 2017. What Defines the "Kingdom" Fungi? Microbiol Spectr 5. https://doi.org/10.1128/microbiolspec.FUNK-0044-2017

Sherman, R.M., Salzberg, S.L., 2020. Pan-genomics in the human genome era. Nat Rev Genet 21, 243–254. https://doi.org/10.1038/s41576-020-0210-7

Sibbald, S.J., Eme, L., Archibald, J.M., Roger, A.J., 2020. Lateral Gene Transfer Mechanisms and Pan-genomes in Eukaryotes. Trends in Parasitology 36, 927–941. https://doi.org/10.1016/j.pt.2020.07.014

Tisdall, J., 2001. Beginning Perl for Bioinformatics: An Introduction to Perl for Biologists. O'Reilly Media, Inc.

Tiwary, B.K., 2020. Chapter 3 - Evolutionary pan-genomics and applications, in: Barh, D., Soares, S., Tiwari, S., Azevedo, V. (Eds.), Pan-Genomics: Applications, Challenges, and Future Prospects. Academic Press, pp. 65–80. https://doi.org/10.1016/B978-0-12-817076-2.00003-2

Vernikos, G., Medini, D., Riley, D.R., Tettelin, H., 2015. Ten years of pan-genome analyses. Current Opinion in Microbiology, Host–microbe interactions: bacteria • Genomics 23, 148–154. https://doi.org/10.1016/j.mib.2014.11.016

Vyas, A., Freitas, A.V., Ralston, Z.A., Tang, Z., 2021. Fission Yeast Schizosaccharomyces pombe: A Unicellular "Micromammal" Model Organism. Curr Protoc 1, e151. https://doi.org/10.1002/cpz1.151

Wapinski, I., Pfeffer, A., Friedman, N., Regev, A., 2007. Natural history and evolutionary principles of gene duplication in fungi. Nature 449, 54–61. https://doi.org/10.1038/nature06107

Zhou, L.-W., May, T.W., 2023. Fungal taxonomy: current status and research agendas for the interdisciplinary and globalisation era. Mycology 14, 52–59. https://doi.org/10.1080/21501203.2022.2103194

7. Appendix

Table 14. List of Pezizomycotina fingerprint proteins

	Fingerprints	NCBI_Description	COG category
1	XP_031039166.2	uncharacterized protein FOBCDRAFT_33666	A
2	XP_054559830.1	uncharacterized protein FOBCDRAFT_290388	A
3	XP_054560205.1	uncharacterized protein FOBCDRAFT_30331	D
4	XP_031044261.2	uncharacterized protein FOBCDRAFT_200621	DP

5	XP_031039445.2	pyridoxal phosphate-dependent transferase	E
6	XP_031040306.1	dihydroxy-acid/6-phosphogluconate dehydratase	E
7	XP_031036147.2	CRAL-TRIO domain-containing protein	I
8	XP_031049849.2	acyl transferase/acyl hydrolase/lysophospholipase	I
9	XP_031039053.2	uncharacterized protein FOBCDRAFT_239710	J
10	XP_031043737.2	uncharacterized protein FOBCDRAFT_212825	J
11	XP_031041560.2	uncharacterized protein FOBCDRAFT_283123	K
12	XP_031042662.1	uncharacterized protein FOBCDRAFT_4138	K
13	XP_031042835.2	fungal-specific transcription factor domain-containing protein	K
14	XP_031048172.2	fungal-specific transcription factor domain-containing protein	K
15	XP_059463851.1	fungal-specific transcription factor domain-containing protein	K
16	XP_059464956.1	uncharacterized protein FOBCDRAFT_293587	K
17	XP_031041616.3	Zinc/iron permease	P
18	XP_031045430.3	uncharacterized protein FOBCDRAFT_220386	Q
19	XP_031029019.2	uncharacterized protein FOBCDRAFT_122967	S
20	XP_031033770.2	uncharacterized protein FOBCDRAFT_231045	S
21	XP_031034276.2	uncharacterized protein FOBCDRAFT_143991	S
22	XP_031035257.1	uncharacterized protein FOBCDRAFT_49653	S
23	XP_031035308.3	uncharacterized protein FOBCDRAFT_242802	S
24	XP_031035325.3	uncharacterized protein FOBCDRAFT_229918	S
25	XP_031036038.3	uncharacterized protein FOBCDRAFT_296934	S
26	XP_031037574.1	uncharacterized protein FOBCDRAFT_320700	S
27	XP_031037581.2	uncharacterized protein FOBCDRAFT_203131	S
28	XP_031037863.2	uncharacterized protein FOBCDRAFT_225737	S
29	XP_031037946.2	cytochrome oxidase c assembly-domain-containing protein	S
30	XP_031038133.2	uncharacterized protein FOBCDRAFT_40438	S
31	XP_031038258.3	uncharacterized protein FOBCDRAFT_137911, partial	S
32	XP_031038397.1	uncharacterized protein FOBCDRAFT_184467	S
33	XP_031038958.3	uncharacterized protein FOBCDRAFT_136206	S
34	XP_031038987.2	uncharacterized protein FOBCDRAFT_33288	S
35	XP_031039176.2	uncharacterized protein FOBCDRAFT_33683	S
36	XP_031039358.2	uncharacterized protein FOBCDRAFT_34254	S
37	XP_031039434.2	uncharacterized protein FOBCDRAFT_319372	S
38	XP_031039572.2	uncharacterized protein FOBCDRAFT_293116	S
39	XP_031039791.2	uncharacterized protein FOBCDRAFT_223477	S
40	XP_031039825.2	uncharacterized protein FOBCDRAFT_35715	S
41	XP_031040318.2	bactericidal permeability-increasing protein	S
42	XP_031041411.1	uncharacterized protein FOBCDRAFT_430	S
43	XP_031041475.2	uncharacterized protein FOBCDRAFT_572	S
44	XP_031042189.2	uncharacterized protein FOBCDRAFT_2714	S
45	XP_031042439.1	uncharacterized protein FOBCDRAFT_3429	S
46	XP_031042527.2	uncharacterized protein FOBCDRAFT_267212	S
47	XP_031042556.2	uncharacterized protein FOBCDRAFT_3798	S
48	XP_031042946.2	uncharacterized protein FOBCDRAFT_235507	S
49	XP_031042972.2	uncharacterized protein FOBCDRAFT_246911	S
50	XP_031042980.2	uncharacterized protein FOBCDRAFT_255694	S
51	XP_031044294.1	uncharacterized protein FOBCDRAFT_31961	S
52	XP_031044440.2	uncharacterized protein FOBCDRAFT_292163	S
53	XP_031045120.2	uncharacterized protein FOBCDRAFT_29286	S
54	XP_031045207.2	major facilitator superfamily domain-containing protein	S
55	XP_031045494.2	uncharacterized protein FOBCDRAFT_28272	S
56	XP_031045835.2	uncharacterized protein FOBCDRAFT_27287	S

57	XP_031046599.3	uncharacterized protein FOBCDRAFT_270576	S
58	XP_031046847.3	uncharacterized protein FOBCDRAFT_198320	S
59	XP_031046995.2	uncharacterized protein FOBCDRAFT_270789	S
60	XP_031047044.2	uncharacterized protein FOBCDRAFT_217408	S
61	XP_031047343.2	uncharacterized protein FOBCDRAFT_21964	S
62	XP_031047835.2	uncharacterized protein FOBCDRAFT_317158	S
63	XP_031048106.2	uncharacterized protein FOBCDRAFT_218701	S
64	XP_031048285.2	uncharacterized protein FOBCDRAFT_317338	S
65	XP_031048300.2	uncharacterized protein FOBCDRAFT_218929	S
66	XP_031048305.2	uncharacterized protein FOBCDRAFT_25002	S
67	XP_031048494.2	uncharacterized protein FOBCDRAFT_219108	S
68	XP_031048497.1	uncharacterized protein FOBCDRAFT_25443	S
69	XP_031048578.2	uncharacterized protein FOBCDRAFT_290718	S
70	XP_031049602.3	uncharacterized protein FOBCDRAFT_287842	S
71	XP_031049608.1	uncharacterized protein FOBCDRAFT_214190	S
72	XP_031049845.2	histidine phosphatase superfamily	S
73	XP_031050301.2	uncharacterized protein FOBCDRAFT_15244	S
74	XP_031050615.2	uncharacterized protein FOBCDRAFT_16104	S
75	XP_031050650.2	uncharacterized protein FOBCDRAFT_215256	S
76	XP_031051020.1	HAD-like domain-containing protein	S
77	XP_031051038.2	fungal-specific transcription factor domain-containing protein	S
78	XP_031051253.1	uncharacterized protein FOBCDRAFT_215956	S
79	XP_054558655.2	XRCC4-like factor-domain-containing protein	S
80	XP_054558769.1	uncharacterized protein FOBCDRAFT_3742	S
81	XP_054558822.1	uncharacterized protein FOBCDRAFT_5223	S
82	XP_054558828.1	uncharacterized protein FOBCDRAFT_5472	S
83	XP_054558934.1	uncharacterized protein FOBCDRAFT_8465	S
84	XP_054559178.1	uncharacterized protein FOBCDRAFT_287776	S
85	XP_054559779.1	uncharacterized protein FOBCDRAFT_22506	S
86	XP_054559845.1	uncharacterized protein FOBCDRAFT_24501	S
87	XP_054559852.2	NmrA-like family-domain-containing protein	S
88	XP_054560176.1	uncharacterized protein FOBCDRAFT_29396	S
89	XP_054560193.1	uncharacterized protein FOBCDRAFT_132082	S
90	XP_054560226.1	uncharacterized protein FOBCDRAFT_31029	S
91	XP_054560288.1	uncharacterized protein FOBCDRAFT_32351	S
92	XP_054561017.1	uncharacterized protein FOBCDRAFT_275504	S
93	XP_054561019.1	uncharacterized protein FOBCDRAFT_163462	S
94	XP_054561027.2	uncharacterized protein FOBCDRAFT_203092	S
95	XP_054562532.1	uncharacterized protein FOBCDRAFT_278362	S
96	XP_054562794.1	uncharacterized protein FOBCDRAFT_298125	S
97	XP_054562809.2	uncharacterized protein FOBCDRAFT_231152	S
98	XP_059463854.1	uncharacterized protein FOBCDRAFT_4726	S
99	XP_059463928.1	Apc15p protein-domain-containing protein	S
100	XP_059464421.1	Saccharopine dehydrogenase-domain-containing protein	S
101	XP_059464499.1	uncharacterized protein FOBCDRAFT_290642	S
102	XP_059464718.1	uncharacterized protein FOBCDRAFT_249831	S
103	XP_059465182.1	uncharacterized protein FOBCDRAFT_320927	S
104	XP_059465621.1	uncharacterized protein FOBCDRAFT_48500	S
105	XP_059466721.1	uncharacterized protein FOBCDRAFT_268882	S
106	XP_031042333.2	kinase-like domain-containing protein	T
107	XP_031037122.2	uncharacterized protein FOBCDRAFT_185529	U
108	XP_031045034.2	P-loop containing nucleoside triphosphate hydrolase protein	U

109	XP_031042647.1	tubulin binding cofactor A	Z

Table 15. List of Saccharomycotina fingerprint proteins

	Fingerprints	NCBI_Description	COG category
1	NP_014497.1	nucleolar protein required for 60S ribosomal subunit biogenesis Nop8p	A
2	NP_014801.1	mitochondrial 37S ribosomal protein PET123	A
3	NP_014492.1	decapping enzyme Dcp1p	AK
4	NP_012359.1	subunit of the SWI/SNF chromatin remodeling complex Swi3p	B
5	NP_010290.3	phosphoribosylanthranilate isomerase TRP1	E
6	NP_010582.3	mitochondrial ribosomal protein of the large subunit Mhr1p	K
7	NP_012208.1	transcriptional activator Vhr1p	K
8	NP_013133.1	protein involved in nucleosome disassembly and transcription elongation from RNA polymerase II promoter Rsc58p	K
9	NP_013324.1	regulator of ribosomal protein (RP) transcription Ifh1p	K
10	NP_013593.1	protein involved in RNA polymerase II transcription initiation Taf8p	K
11	NP_014574.1	transcription factor (bHLH) involved in interorganelle communication Rtg1p	K
12	NP_010636.1	putative aspartic endopeptidase	O
13	NP_013134.1	putative divalent metal ion transporter SMF3	P
14	NP_010176.1	peripherally bound inner membrane protein of the mitochondrial matrix Mss2p	S
15	NP_011834.1	cytoplasmic protein required for replication of Brome mosaic virus Oca5p	S
16	NP_014885.3	sporulation specific protein that localizes to the spore wall; required for sporulation at a point after meiosis II and during spore wall formation Ssp2p	S
17	NP_013664.1	endoplasmic reticulum (ER) localized integral membrane protein Gsf2p	U
18	NP_014974.3	protein required for normal actin organization and endocytosis Scd5p	U

Table 16. List of Schizosaccharomyces fingerprint proteins

	Fingerprints	NCBI_Description	COG category
1	NP_588373.1	rrm type RNA-binding protein	A
2	NP_593572.1	putative TRAMP complex ATP-dependent RNA helicase	A
3	NP_594243.1	RNA-binding protein Csx1	A
4	NP_596836.2	U1 snRNP-associated protein Usp109	A
5	NP_592937.1	putative histone demethylase SWIRM2	B
6	NP_593646.1	Clr6 histone deacetylase complex subunit Pst2	B
7	NP_594864.1	Clr6 histone deacetylase complex subunit Prw1	B
8	NP_595433.1	SIN3 family co-repressor Pst3	B
9	NP_596495.1	putative NatA N-acetyltransferase complex subunit	B
10	NP_593328.1	putative inorganic diphosphatase	C
11	NP_593651.1	glycerol dehydrogenase Gld1	C
12	NP_594711.1	putative mRNA-processing protein Cox24	C
13	NP_595261.1	putative NADH dehydrogenase	C

14	NP_595639.1	Lid2 complex histone demethylase H3-K36 subunit Jmj3	C
15	NP_588258.1	thioredoxin family protein	CO
16	NP_588422.1	nucleoporin Npp106	D
17	NP_593057.1	putative serine/threonine protein kinase Ppk8	D
18	NP_593889.1	G1/S-specific B-type cyclin Cig2	D
19	NP_595093.1	mitotic inhibitor kinase Mik1	D
20	NP_595210.1	putative poly(A) polymerase Cid11	D
21	NP_001342803.1	putative GTPase-activating protein	DT
22	NP_001343003.1	septin Spn7	DUZ
23	NP_001342763.1	putative hydantoin racemase family protein	E
24	NP_588377.1	gamma-aminobutyric acid/polyamine transporter	E
25	NP_593172.1	putative succinate-semialdehyde dehydrogenase	E
26	NP_593349.2	putative 5'-methylthioribulose-1-phosphate dehydratase	E
27	NP_595830.1	phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphohydrolase His7	E
28	NP_596348.1	amino-acid permease	E
29	NP_596769.2	putative APC amino acid transporter	E
30	NP_596849.1	amino-acid permease	E
31	NP_594250.1	putative phosphoprotein phosphatase	F
32	NP_001018282.1	alpha-1,3-glucan synthase Mok11	G
33	NP_587978.1	alpha-1,4-glucan synthase Mok14	G
34	NP_593152.1	putative 6-phosphofructo-2-kinase	G
35	NP_593196.1	lysophospholipase 1	G
36	NP_593614.1	putative glucose-6-phosphate 1-dehydrogenase	G
37	NP_594024.1	putative phospholipase	G
38	NP_594171.1	putative alpha-1,2-mannosyltransferase Omh6	G
39	NP_594258.1	putative glucosidase	G
40	NP_594975.1	putative trehalose-phosphate synthase Tps2	G
41	NP_595075.1	putative mannan endo-1,6-alpha-mannosidase	G
42	NP_595564.1	putative serine/threonine protein kinase Gsk31	G
43	NP_595619.1	alpha-1,3-glucan synthase Mok12	G
44	NP_596053.1	putative 1,3-beta-glucanosyltransferase Gas2	G
45	NP_588294.1	putative nucleoside 2-deoxyribosyltransferase	H
46	NP_594427.1	decaprenyl diphosphate synthase subunit 2 Dlp1	H
47	NP_595334.1	farnesyl pyrophosphate synthetase	H
48	NP_596437.2	trifunctional formate-tetrahydrofolate ligase/methenyltetrahydrofolate cyclohydrolase/methylenetetrahydrofolate dehydrogenase	H
49	NP_587935.1	putative phosphomethylpyrimidine kinase	HK
50	NP_596524.1	TENA/THI family protein	HK
51	NP_001342934.1	putative inositol polyphosphate phosphatase	I
52	NP_588302.1	putative ribonucleoprotein complex protein	I
53	NP_592859.3	putative GNS1/SUR4 family protein	I
54	NP_594057.1	putative sec14 family protein	I
55	NP_595918.1	putative triglyceride lipase-cholesterol esterase	I
56	NP_593037.1	putative AMP binding enzyme	IQ
57	NP_592958.1	putative autophagy C terminal domain family protein	J
58	NP_593079.1	Clr6 histone deacetylase-associated PHD protein-1 Cph1	J
59	NP_593339.1	translation initiation factor eIF3j (p35)	J
60	NP_593454.1	ribosome biogenesis protein rlp7	J
61	NP_594793.1	putative Noc2-Noc3 complex subunit	J
62	NP_594944.1	putative ski complex-interacting GTPase	J
63	NP_595442.1	translation termination factor Rrf1	J

64	NP_596485.1	putative tRNA pseudouridine synthase Pus2	J
65	NP_001342783.1	putative GTPase-activating protein	K
66	NP_001343126.1	protein erh1	K
67	NP_594500.1	transcription factor Pcr1	K
68	NP_595324.1	putative transcription factor	K
69	NP_595687.1	DNA methyltransferase	K
70	NP_595712.1	putative EST1 family protein	K
71	NP_596183.1	transcription factor Rsv1	K
72	NP_596254.1	transcription initiation factor TFIID subunit taf73	K
73	NP_001342977.1	putative ATP-dependent 3'-5'-directionality DNA helicase fml2	L
74	NP_588204.1	silencing protein Rik1	L
75	NP_588273.1	poly(A) polymerase Cid12	L
76	NP_594246.1	putative ATP-dependent DNA helicase	L
77	NP_595514.1	3'-tRNA-processing endonuclease Trz2	L
78	NP_596275.1	phosphatidylinositol kinase Tor1	L
79	NP_594032.1	1,3-beta-glucan synthase subunit Bgs2	M
80	NP_594766.1	1,3-beta-glucan synthase subunit Bgs3	M
81	NP_595434.1	chitin synthase-like protein 2	M
82	NP_595971.1	1,3-beta-glucan synthase catalytic subunit Bgs1	M
83	NP_001342703.1	putative ubiquitin family protein	O
84	NP_587999.1	ubiquitin hydrolase Ubp1	O
85	NP_588507.1	putative protein disulfide isomerase	O
86	NP_593439.1	SUMO-targeted ubiquitin-protein ligase subunit Rfp2	O
87	NP_593926.1	HECT-type ubiquitin-protein ligase Pub2	O
88	NP_594172.4	putative protein disulfide isomerase	O
89	NP_594377.1	putative MatE family transporter	O
90	NP_594920.2	putative dipeptidyl peptidase	O
91	NP_595348.1	uncharacterized protein SPBC713.09	O
92	NP_595793.1	putative HECT-type ubiquitin-protein ligase Pub3	O
93	NP_595954.2	thioredoxin Trx2	O
94	NP_596652.1	SWI/SNF complex subunit Snf59	O
95	NP_594814.1	putative COP9/signalosome complex subunit 7a	OT
96	NP_594508.1	putative cornichon family protein	OTU
97	NP_001342754.1	ER oxidoreductin Ero1a	OU
98	NP_001342909.1	potassium ion transporter Trk2	P
99	NP_594278.1	putative rRNA processing protein Fyv7	P
100	NP_594759.1	putative P-type ATPase	P
101	NP_594863.1	Ca(2+)/Mn(2+)-transporting P-type ATPase P5 type Cta5	P
102	NP_588242.1	putative AMP-binding dehydrogenase	Q
103	NP_594585.1	ABC transporter Ybt1	Q
104	NP_001342727.1	uncharacterized protein SPBC36.11	S
105	NP_001342742.1	putative 2',3'-cyclic-nucleotide 3'-phosphodiesterase	S
106	NP_001342749.1	meiotically upregulated gene Mug174	S
107	NP_001342820.1	hepatocellular carcinoma-associated antigen 59 domain-containing protein	S
108	NP_001342836.1	vitatmin B6 import carrier Bsu1	S
109	NP_001342867.1	ATR checkpoint kinase regulatory subunit Rad26	S
110	NP_001342874.1	LOW QUALITY PROTEIN: protein Lsd90	S
111	NP_001342981.1	uncharacterized protein SPBC16E9.19	S
112	NP_001342990.1	uncharacterized protein SPBP22H7.04	S
113	NP_001343020.1	bouquet formation protein Bqt1	S
114	NP_001343076.1	uncharacterized protein SPAC688.16	S

115	NP_001343077.1	Ino80 complex subunit Iec5	S
116	NP_001343080.1	uncharacterized protein SPAC23D3.16	S
117	NP_001343084.1	uncharacterized protein SPBC29A10.17	S
118	NP_001343089.1	uncharacterized protein SPCC757.15	S
119	NP_001343100.1	protein tam1	S
120	NP_001343102.1	uncharacterized protein SPAC17A5.19	S
121	NP_001343110.1	protein new2	S
122	NP_001343112.1	ribonucleotide reductase (RNR) inhibitor family	S
123	NP_001343122.1	uncharacterized protein SPAC4H3.16	S
124	NP_001343149.1	uncharacterized protein SPBC30B4.09	S
125	NP_001343159.1	protein tam10	S
126	NP_001343161.1	putative TOM complex subunit Tom5	S
127	NP_001343177.1	protein tam14	S
128	NP_001343183.1	uncharacterized protein SPCPB16A4.07	S
129	NP_001343187.1	GTPAse-interacting protein	S
130	NP_587686.1	putative transporter	S
131	NP_587773.1	uncharacterized protein SPCC736.02	S
132	NP_587793.1	bouquet formation protein Bqt3	S
133	NP_587843.1	calnexin independence factor Cif1	S
134	NP_587873.1	switch-activating protein Sap1	S
135	NP_587911.1	putative Ino80 complex subunit	S
136	NP_587939.1	uncharacterized protein SPCC18B5.09c	S
137	NP_587970.1	uncharacterized protein SPCC1393.12	S
138	NP_587974.1	ACN9 family protein	S
139	NP_587975.2	uncharacterized protein SPCC2H8.05c	S
140	NP_587990.1	uncharacterized protein SPCC16A11.03c	S
141	NP_588002.1	uncharacterized protein SPCC16A11.15c	S
142	NP_588006.1	uncharacterized protein SPCC24B10.03	S
143	NP_588007.1	uncharacterized protein SPCC24B10.04	S
144	NP_588009.1	uncharacterized protein SPCC24B10.06	S
145	NP_588019.2	uncharacterized protein SPCC24B10.16c	S
146	NP_588032.1	uncharacterized protein SPCC1795.12c	S
147	NP_588075.1	medial ring protein Mid1	S
148	NP_588120.1	protein mug117	S
149	NP_588125.2	putative pre-ribosomal factor	S
150	NP_588210.1	telomere maintenance protein Ccq1	S
151	NP_588220.1	arrestin/PY protein 2	S
152	NP_588232.1	TORC1 subunit Tco89	S
153	NP_588290.1	uncharacterized protein SPCC191.01	S
154	NP_588402.1	uncharacterized protein SPCC4F11.03c	S
155	NP_588443.1	putative pyridoxamine/pyridoxine/pyridoxal transmembrane transporter	S
156	NP_588489.1	uncharacterized protein SPCC1919.07	S
157	NP_588544.1	protein mug9	S
158	NP_592868.1	uncharacterized protein SPAC1F5.05c	S
159	NP_592901.1	uncharacterized protein SPAC630.06c	S
160	NP_592902.1	uncharacterized protein SPAC630.07c	S
161	NP_592944.1	uncharacterized protein SPAC24H6.08	S
162	NP_593113.2	protein mug35	S
163	NP_593156.1	uncharacterized protein SPAC821.03c	S
164	NP_593198.1	protein meu31	S
165	NP_593284.1	uncharacterized protein SPAC1565.05	S

166	NP_593309.2	putative Rho guanine nucleotide exchange factor	S
167	NP_593493.1	bouquet formation protein Bqt2	S
168	NP_593519.2	uncharacterized protein SPAPB1A10.05	S
169	NP_593528.1	putative F-box protein	S
170	NP_593550.1	uncharacterized protein SPAC3H1.08c	S
171	NP_593562.1	uncharacterized protein SPAC9G1.07	S
172	NP_593596.1	uncharacterized protein SPAC607.07c	S
173	NP_593599.1	sporulation protein Spo3	S
174	NP_593618.1	uncharacterized protein SPAC25A8.02	S
175	NP_593628.1	putative phosphatase activator	S
176	NP_593675.1	uncharacterized protein SPAC18G6.12c	S
177	NP_593676.1	uncharacterized protein SPAC18G6.13	S
178	NP_593679.1	F-box protein Pof14	S
179	NP_593688.1	putative histone chaperone Chz1	S
180	NP_593735.1	uncharacterized protein SPAC17G8.12	S
181	NP_593758.1	HbrB family protein	S
182	NP_593805.1	uncharacterized protein SPAC23H3.15c	S
183	NP_593809.1	protein meu29	S
184	NP_593857.1	SH3 domain-containing protein	S
185	NP_593878.1	putative mediator complex subunit Med11	S
186	NP_593945.1	UBA domain protein Ucp6	S
187	NP_593954.1	protein mug130	S
188	NP_593967.1	M-phase phosphoprotein 6 family protein	S
189	NP_594007.2	tetraspan protein Dni1	S
190	NP_594029.1	uncharacterized protein SPAC24C9.04	S
191	NP_594043.2	putative Pep5/Vps11-like zinc finger domain-containing protein	S
192	NP_594045.2	DASH complex subunit Dad1	S
193	NP_594051.1	uncharacterized protein SPAC589.03c	S
194	NP_594070.1	uncharacterized protein SPAC688.12c	S
195	NP_594113.1	sporulation protein Spo7	S
196	NP_594124.1	uncharacterized protein SPAC6G9.15c	S
197	NP_594125.2	uncharacterized protein SPAC6G9.16c	S
198	NP_594127.1	uncharacterized protein SPAPB1E7.01c	S
199	NP_594140.1	putative Rab GTPase binding protein upregulated in meiosis II	S
200	NP_594217.1	protein mde1	S
201	NP_594231.1	transcription factor Deb1/Rdp1	S
202	NP_594241.1	uncharacterized protein SPAC17A2.07c	S
203	NP_594260.1	protein slt1	S
204	NP_594280.1	protein mug129	S
205	NP_594287.1	DASH complex subunit Spc34	S
206	NP_594294.2	uncharacterized protein SPAC15A10.07	S
207	NP_594338.1	Swr1 complex subunit Swc3	S
208	NP_594342.3	uncharacterized protein SPAC4H3.06	S
209	NP_594350.1	uncharacterized protein SPAC4H3.14c	S
210	NP_594404.1	uncharacterized protein SPAC27E2.11c	S
211	NP_594408.1	galactose residue biosynthesis protein Pvg2	S
212	NP_594428.1	Pot1-associated protein Poz1	S
213	NP_594435.1	putative serine palmitoyl transferase A subunit	S
214	NP_594437.1	putative ubiquitin-protein ligase E3	S
215	NP_594513.1	uncharacterized protein SPAC2C4.10c	S
216	NP_594545.2	U1 snRNP-associated protein Usp108	S

217	NP_594591.1	uncharacterized protein SPAC17C9.15c	S
218	NP_594601.1	mediator complex subunit Pmc3/Med27	S
219	NP_594611.1	anaphase-promoting complex subunit Apc14	S
220	NP_594634.1	uncharacterized protein SPAC12B10.02c	S
221	NP_594642.1	uncharacterized protein SPAC12B10.10	S
222	NP_594655.1	uncharacterized protein SPAC30C2.03	S
223	NP_594662.1	uncharacterized protein SPAC144.01	S
224	NP_594725.1	PvGal biosynthesis protein Pvg5	S
225	NP_594735.1	small histone ubiquitination factor Shf1	S
226	NP_594784.2	uncharacterized protein SPAPB8E5.10	S
227	NP_594812.1	uncharacterized protein SPAC1952.10c	S
228	NP_594817.1	meiotic recombination protein Rec24	S
229	NP_594870.1	protein mug108	S
230	NP_594933.1	Smc5-6 complex non-SMC subunit Nse6	S
231	NP_594943.1	heat shock protein Hsp9	S
232	NP_594986.3	uncharacterized protein SPAC29B12.08	S
233	NP_595087.2	uncharacterized protein SPBC660.08	S
234	NP_595096.1	uncharacterized protein SPBC660.17c	S
235	NP_595103.1	putative cell cycle transcriptional repressor Whi5	S
236	NP_595182.1	putative nuclear membrane organization protein Apq12	S
237	NP_595215.1	RecA family ATPase Rlp1	S
238	NP_595224.1	protein Cut12	S
239	NP_595262.1	putative respiratory chain complex assembly protein Cbp6	S
240	NP_595391.1	uncharacterized protein SPBC146.02	S
241	NP_595400.2	protein mug97	S
242	NP_595474.1	putative tetraspan protein Dni2	S
243	NP_595502.1	uncharacterized protein SPBC651.04	S
244	NP_595504.3	protein mug166	S
245	NP_595505.1	uncharacterized protein SPBC651.07	S
246	NP_595508.1	Smc5-6 complex non-SMC subunit Nse5	S
247	NP_595513.1	But2 family protein But2	S
248	NP_595573.1	uncharacterized protein SPBC8D2.11	S
249	NP_595600.1	putative beta-glucan synthesis-associated protein	S
250	NP_595658.1	protein meu25	S
251	NP_595679.1	uncharacterized protein SPBC21B10.08c	S
252	NP_595705.1	transcriptional activator MBF subunit Rep2	S
253	NP_595723.1	putative DNA-directed RNA polymerase I complex subunit Rpa34	S
254	NP_595783.2	CCAAT-binding factor complex subunit Php4	S
255	NP_595811.1	uncharacterized protein SPBC1A4.11c	S
256	NP_595853.1	uncharacterized protein SPBC18E5.07	S
257	NP_595887.1	meiotic recombination protein Rec15	S
258	NP_595902.1	uncharacterized protein SPBC17G9.13c	S
259	NP_595993.1	Tor complex Tor2-interacting protein 1	S
260	NP_596005.1	uncharacterized protein SPBC21D10.08c	S
261	NP_596007.1	agglutination protein Map4	S
262	NP_596040.1	GRIP domain protein	S
263	NP_596142.1	uncharacterized protein SPBC685.08	S
264	NP_596161.1	uncharacterized protein SPBC19C7.05	S
265	NP_596188.1	uncharacterized protein SPBP4H10.14c	S
266	NP_596251.1	putative Mam33 family protein	S
267	NP_596261.1	mitotic-spindle disanchored Msd1	S

268	NP_596308.1	uncharacterized protein SPBC405.05	S
269	NP_596395.1	uncharacterized protein SPBC11C11.06c	S
270	NP_596439.1	protein mug110	S
271	NP_596449.1	DNA-directed RNA polymerase I complex subunit Ker1	S
272	NP_596456.2	putative rRNA-processing Rrp12-like protein	S
273	NP_596482.1	uncharacterized protein SPBC887.08	S
274	NP_596508.1	putative Rnase P/MRP subunit Pop7	S
275	NP_596517.2	putative U3 snoRNP-associated protein Utp16	S
276	NP_596597.1	uncharacterized protein SPBC21C3.17c	S
277	NP_596600.1	C2 domain protein Git1	S
278	NP_596628.1	uncharacterized protein SPBC1604.12	S
279	NP_596637.1	uncharacterized protein SPBC1604.03c	S
280	NP_596695.1	sporulation protein Meu14	S
281	NP_596723.1	uncharacterized protein SPBC1861.06c	S
282	NP_596728.1	putative TRAPP complex subunit Trs65	S
283	NP_596755.1	uncharacterized protein SPBC16G5.06	S
284	NP_596762.1	protein ptf2	S
285	NP_596782.1	protein nrm1	S
286	NP_001342764.1	serine/threonine protein phosphatase Sds21	T
287	NP_001342881.1	serine/threonine protein kinase Ppk21	T
288	NP_588283.1	protein kinase Mug27/Slk1	T
289	NP_588555.1	Ark1/Prk1 family protein kinase Ppk38	T
290	NP_593132.1	calcium-binding protein	T
291	NP_593184.1	serine/threonine protein kinase Hhp2	T
292	NP_593404.1	putative serine/threonine protein kinase Ppk9	T
293	NP_594840.1	serine/threonine protein kinase Sck2	T
294	NP_594856.2	MAP kinase kinase kinase Win1	T
295	NP_595231.3	putative oxysterol binding protein	T
296	NP_595486.1	putative serine/threonine protein kinase Ppk31	T
297	NP_595581.1	serine/threonine protein kinase Mde3	T
298	NP_596099.1	putative chitin synthase regulatory factor Chr1	T
299	NP_596726.1	putative serine/threonine protein kinase Ppk22	T
300	NP_593510.1	putative uracil phosphoribosyltransferase	TZ
301	NP_587995.1	putative sorting nexin Atg20	U
302	NP_588067.1	putative GTPase-activating protein Gyp2	U
303	NP_593008.1	adaptin family protein	U
304	NP_593524.1	GTPase Ypt71	U
305	NP_593843.1	WD repeat-containing protein	U
306	NP_594041.1	uncharacterized protein SPAC16A10.01	U
307	NP_595311.1	inositol-1,4,5-trisphosphate 5-phosphatase 2	U
308	NP_595884.1	putative autophagy associated protein	U
309	NP_595983.2	putative translocon subunit Sec61-like protein	U
310	NP_596319.1	putative COPII cargo receptor subunit Sec23b	U
311	NP_596800.2	vacuolar protein sorting-associated protein	U
312	NP_001342876.1	putative septin Spn6	Z
313	NP_587785.1	microtubule-associated protein Dis1	Z
314	NP_588046.1	formin For3	Z
315	NP_593816.1	Myo3-like myosin II heavy chain myp2	Z
316	NP_594040.1	septin Spn5	Z
317	NP_594877.1	myosin I light chain Cam2	Z
318	NP_595633.2	UBA/EH/EF hand domain protein Ucp8	Z
319	NP_596233.1	myosin type V	Z

Table 17. List of Candida fingerprint proteins

	Fingerprints	NCBI_Description	COG category
1	XP_717177.2	Cdl1p	A
2	XP_723187.1	uncharacterized protein CAALFM_C305060WA	A
3	XP_711045.2	uncharacterized protein CAALFM_CR07780WA	C
4	XP_711412.2	Gap6p	E
5	XP_719453.2	Hnm1p	E
6	XP_720611.1	Gap1p	E
7	XP_722739.2	Aat22p	E
8	XP_710800.2	Rnr3p	F
9	XP_710275.2	Mnt3p	G
10	XP_722202.1	Kis2p	G
11	XP_714291.1	uncharacterized protein CAALFM_C205990CA	I
12	XP_715205.1	Faa2-3p	I
13	XP_716979.1	uncharacterized protein CAALFM_C703040WA	I
14	XP_715141.2	uncharacterized protein CAALFM_CR03660CA	J
15	XP_722548.1	uncharacterized protein CAALFM_C201910WA	J
16	XP_716951.2	uncharacterized protein CAALFM_C703270WA	JO
17	XP_711489.1	Efh1p	K
18	XP_713154.1	uncharacterized protein CAALFM_C207370WA	K
19	XP_716760.1	Hms1p	K
20	XP_717795.1	telomerase subunit	K
21	XP_719121.1	Rbf1p	K
22	XP_721546.2	uncharacterized protein CAALFM_C103460CA	K
23	XP_721779.2	uncharacterized protein CAALFM_C301430WA	K
24	XP_721337.2	uncharacterized protein CAALFM_C702160WA	KL
25	XP_717760.2	Chs8p	M
26	XP_714833.1	Prb1p	O
27	XP_716247.1	uncharacterized protein CAALFM_CR08420WA	O
28	XP_719311.1	putative dolichyl-phosphate-mannose--protein mannosyltransferase	O
29	XP_712626.1	uncharacterized protein CAALFM_C400700CA	P
30	XP_019330611.1	uncharacterized protein CAALFM_C101610CA	S
31	XP_019330634.1	uncharacterized protein CAALFM_C103840WA	S
32	XP_019330741.1	uncharacterized protein CAALFM_C200730CA	S
33	XP_019330861.1	uncharacterized protein CAALFM_C304330CA	S
34	XP_019331025.1	uncharacterized protein CAALFM_C701070CA	S
35	XP_441044.2	uncharacterized protein CAALFM_CR06060WA	S
36	XP_710133.2	uncharacterized protein CAALFM_C500550CA	S
37	XP_710173.1	uncharacterized protein CAALFM_CR07120CA	S
38	XP_710208.2	uncharacterized protein CAALFM_CR07680CA	S
39	XP_710296.1	Pex22p	S
40	XP_710316.2	uncharacterized protein CAALFM_CR05210WA	S
41	XP_710373.2	uncharacterized protein CAALFM_C206740WA	S
42	XP_710411.2	uncharacterized protein CAALFM_C104350CA	S
43	XP_710478.2	uncharacterized protein CAALFM_C200090WA	S
44	XP_710647.2	uncharacterized protein CAALFM_C300890CA	S
45	XP_710792.2	uncharacterized protein CAALFM_C112540WA	S
46	XP_710933.1	uncharacterized protein CAALFM_C307820WA	S
47	XP_711053.1	uncharacterized protein CAALFM_CR07850WA	S

48	XP_711317.2	uncharacterized protein CAALFM_C600800CA	S
49	XP_711332.2	uncharacterized protein CAALFM_C600690WA	S
50	XP_711352.2	uncharacterized protein CAALFM_C501530CA	S
51	XP_711411.1	uncharacterized protein CAALFM_C503490CA	S
52	XP_711701.2	uncharacterized protein CAALFM_CR05770WA	S
53	XP_711731.2	uncharacterized protein CAALFM_C107560WA	S
54	XP_711829.1	uncharacterized protein CAALFM_C206440CA	S
55	XP_712095.2	uncharacterized protein CAALFM_C604240WA	S
56	XP_712243.1	uncharacterized protein CAALFM_C104620WA	S
57	XP_712297.2	uncharacterized protein CAALFM_C600200CA	S
58	XP_712319.1	uncharacterized protein CAALFM_C600400CA	S
59	XP_712342.2	uncharacterized protein CAALFM_CR07340CA	S
60	XP_712509.2	uncharacterized protein CAALFM_C602280WA	S
61	XP_712643.1	uncharacterized protein CAALFM_C400530CA	S
62	XP_712740.1	uncharacterized protein CAALFM_C303770CA	S
63	XP_712924.1	uncharacterized protein CAALFM_C204520CA	S
64	XP_712939.1	uncharacterized protein CAALFM_C204400WA	S
65	XP_712969.1	uncharacterized protein CAALFM_C101800WA	S
66	XP_713130.1	uncharacterized protein CAALFM_C207540WA	S
67	XP_713214.1	uncharacterized protein CAALFM_CR03410WA	S
68	XP_713316.1	Pga7p	S
69	XP_713430.1	uncharacterized protein CAALFM_CR02270CA	S
70	XP_713431.2	uncharacterized protein CAALFM_CR02260CA	S
71	XP_713506.2	uncharacterized protein CAALFM_CR06290CA	S
72	XP_713794.2	uncharacterized protein CAALFM_C111430WA	S
73	XP_713873.1	uncharacterized protein CAALFM_C107400CA	S
74	XP_713899.2	uncharacterized protein CAALFM_C107130CA	S
75	XP_713909.1	uncharacterized protein CAALFM_C111120CA	S
76	XP_714303.1	uncharacterized protein CAALFM_C600930CA	S
77	XP_714311.2	uncharacterized protein CAALFM_C601000CA	S
78	XP_714394.2	uncharacterized protein CAALFM_C210300CA	S
79	XP_714451.2	uncharacterized protein CAALFM_C208690CA	S
80	XP_714556.2	uncharacterized protein CAALFM_C209690CA	S
81	XP_714581.2	uncharacterized protein CAALFM_C502060WA	S
82	XP_714922.2	uncharacterized protein CAALFM_C205330CA	S
83	XP_715010.1	uncharacterized protein CAALFM_C114020WA	S
84	XP_715026.1	uncharacterized protein CAALFM_C114180WA	S
85	XP_715381.1	uncharacterized protein CAALFM_C202660WA	S
86	XP_715750.2	uncharacterized protein CAALFM_CR02600WA	S
87	XP_715784.1	uncharacterized protein CAALFM_CR02960WA	S
88	XP_715848.2	uncharacterized protein CAALFM_C105520WA	S
89	XP_715861.2	uncharacterized protein CAALFM_C105390CA	S
90	XP_716071.1	uncharacterized protein CAALFM_C306260CA	S
91	XP_716104.1	uncharacterized protein CAALFM_C306540CA	S
92	XP_716253.1	uncharacterized protein CAALFM_CR08450CA	S
93	XP_716274.1	uncharacterized protein CAALFM_CR08620CA	S
94	XP_716285.1	uncharacterized protein CAALFM_CR08710WA	S
95	XP_716323.1	uncharacterized protein CAALFM_C112110CA	S
96	XP_716547.1	uncharacterized protein CAALFM_C405580CA	S
97	XP_716554.2	uncharacterized protein CAALFM_C405610CA	S
98	XP_716585.2	uncharacterized protein CAALFM_C405900CA	S
99	XP_716748.1	Frp6p	S

100	XP_716767.2	uncharacterized protein CAALFM_C500610CA	S
101	XP_716774.2	Aga1p	S
102	XP_717193.2	Spc34p	S
103	XP_717223.1	Goa1p	S
104	XP_717249.1	Fgr3p	S
105	XP_717615.1	uncharacterized protein CAALFM_C108050WA	S
106	XP_717631.1	Bul1p	S
107	XP_717814.1	uncharacterized protein CAALFM_C300270CA	S
108	XP_717936.2	uncharacterized protein CAALFM_CR01090WA	S
109	XP_717956.1	uncharacterized protein CAALFM_CR00880WA	S
110	XP_718138.2	Cta9p	S
111	XP_718146.1	uncharacterized protein CAALFM_C101340CA	S
112	XP_718199.1	Csp37p	S
113	XP_718473.2	Pga45p	S
114	XP_718474.1	uncharacterized protein CAALFM_C105950CA	S
115	XP_718481.2	uncharacterized protein CAALFM_C105900WA	S
116	XP_718531.2	uncharacterized protein CAALFM_C301110CA	S
117	XP_718651.1	Mms22p	S
118	XP_718851.1	uncharacterized protein CAALFM_C106500WA	S
119	XP_719016.2	uncharacterized protein CAALFM_C100540CA	S
120	XP_719083.1	uncharacterized protein CAALFM_C602490CA	S
121	XP_719503.2	Fgr15p	S
122	XP_719830.1	uncharacterized protein CAALFM_C307590WA	S
123	XP_719963.2	uncharacterized protein CAALFM_C302750WA	S
124	XP_719981.2	uncharacterized protein CAALFM_C302910WA	S
125	XP_720010.1	uncharacterized protein CAALFM_C303160CA	S
126	XP_720047.2	uncharacterized protein CAALFM_C303540WA	S
127	XP_720352.1	uncharacterized protein CAALFM_C700590WA	S
128	XP_720375.1	Dfi1p	S
129	XP_720376.1	uncharacterized protein CAALFM_C700350CA	S
130	XP_720556.2	uncharacterized protein CAALFM_C502390CA	S
131	XP_720562.1	glucose-induced degradation complex subunit	S
132	XP_720637.1	uncharacterized protein CAALFM_C503040WA	S
133	XP_720937.1	uncharacterized protein CAALFM_C204080WA	S
134	XP_721402.1	uncharacterized protein CAALFM_C701590CA	S
135	XP_721507.2	uncharacterized protein CAALFM_C103170CA	S
136	XP_722145.2	Op4p	S
137	XP_722194.1	ssDNA endodeoxyribonuclease	S
138	XP_722223.1	uncharacterized protein CAALFM_C113820CA	S
139	XP_722530.2	Hap42p	S
140	XP_722533.2	kinetochore-associated Ndc80 complex subunit	S
141	XP_722563.2	uncharacterized protein CAALFM_C202040WA	S
142	XP_722587.1	uncharacterized protein CAALFM_C202260WA	S
143	XP_722662.2	Rfx2p	S
144	XP_722667.1	uncharacterized protein CAALFM_C401970WA	S
145	XP_722714.1	uncharacterized protein CAALFM_C401430CA	S
146	XP_722776.2	uncharacterized protein CAALFM_C400840WA	S
147	XP_723116.1	Ras2p	S
148	XP_723186.1	Try4p	S
149	XP_723421.1	uncharacterized protein CAALFM_C108760WA	S
150	XP_723575.2	uncharacterized protein CAALFM_C110230CA	S
151	XP_711842.2	uncharacterized protein CAALFM_C206600WA	T

152	XP_718801.2	uncharacterized protein CAALFM_C106970CA	T
153	XP_710621.1	uncharacterized protein CAALFM_C204730WA	U
154	XP_711567.2	Rvs162p	U
155	XP_721394.1	uncharacterized protein CAALFM_C701660CA	U
156	XP_721528.1	uncharacterized protein CAALFM_C103330CA	U
157	XP_718049.2	FG-nucleoporin	UY

Table 18. List of Saccharomyces fingerprint proteins

	Fingerprints	NCBI_Description	COG category
1	NP_009770.3	Ame1p	A
2	NP_010350.3	Rrg1p	A
3	NP_010470.1	Atc1p	A
4	NP_011017.3	Ies5p	A
5	NP_011685.3	pseudouridine synthase PUS6	A
6	NP_011695.1	Okp1p	A
7	NP_012708.1	Yra2p	A
8	NP_013312.2	Atg38p	A
9	NP_013589.1	Nab6p	A
10	NP_015382.1	Brr1p	A
11	NP_010637.4	Atp22p	C
12	NP_013586.1	NADH-ubiquinone reductase (H(+)-translocating) NDI1	C
13	NP_015445.1	B-type cyclin CLB5	D
14	NP_011579.1	Vht1p	G
15	NP_011598.3	mitochondrial 37S ribosomal protein YmS-A	J
16	NP_012792.1	She2p	J
17	NP_010440.1	DNA-directed RNA polymerase I subunit RPA14	K
18	NP_010481.3	RNA-processing protein REF2	K
19	NP_010596.1	Sum1p	K
20	NP_012037.1	Thp2p	K
21	NP_013153.1	Ies3p	K
22	NP_013196.1	Ioc2p	K
23	NP_013489.1	Swc7p	K
24	NP_013997.1	Rrn9p	K
25	NP_014200.1	Gcr2p	K
26	NP_014709.1	Msa1p	K
27	NP_116605.1	Swp82p	K
28	NP_010363.1	Shu2p	L
29	NP_010395.1	replication fork barrier binding protein FOB1	L
30	NP_012134.1	Csm2p	L
31	NP_012362.1	Rfa3p	L
32	NP_009600.2	Tcm62p	O
33	NP_009995.1	peptidylprolyl isomerase family protein CPR4	O
34	NP_013314.1	putative glycosylase	O
35	NP_014553.1	protein disulfide isomerase MPD2	O
36	NP_015251.1	putative AAA family ATPase YTA6	O
37	NP_013770.1	H(+)-transporting V0 sector ATPase subunit a	P
38	NP_009352.2	gamma-tubulin complex subunit SPC72	S
39	NP_009375.1	DNA-binding protein SAW1	S
40	NP_009724.1	Ubs1p	S
41	NP_010005.1	Ahc2p	S

42	NP_010101.1	Ait1p	S
43	NP_010162.1	Exp1p	S
44	NP_010210.1	Ahk1p	S
45	NP_010341.3	Emc10p	S
46	NP_010472.3	Snd1p	S
47	NP_010949.4	Fir1p	S
48	NP_011074.3	cohesin-loading factor complex subunit SCC4	S
49	NP_011479.1	Vir1p	S
50	NP_011530.1	uncharacterized protein YGR016W	S
51	NP_011646.1	uncharacterized protein YGR130C	S
52	NP_012130.1	Om45p	S
53	NP_012209.1	uncharacterized protein YIL055C	S
54	NP_012388.1	Smt1p	S
55	NP_012518.2	Tph3p	S
56	NP_012848.1	Aan1p	S
57	NP_012972.3	Pln1p	S
58	NP_013683.1	Usa1p	S
59	NP_013689.1	Smc5-Smc6 complex subunit NSE5	S
60	NP_013767.1	Far3p	S
61	NP_013864.1	Fdo1p	S
62	NP_013883.1	Cvm1p	S
63	NP_013984.1	Pet111p	S
64	NP_014027.1	sphingosine N-acyltransferase subunit LIP1	S
65	NP_014090.2	Stb1p	S
66	NP_014104.1	Mrx6p	S
67	NP_014240.1	Asi2p	S
68	NP_014443.3	Pet494p	S
69	NP_014922.3	Rfm1p	S
70	NP_015167.1	Aim44p	S
71	NP_015474.1	uncharacterized protein YPR148C	S
72	NP_061491.1	uncharacterized protein YBR085C-A	S
73	NP_116663.1	Far7p	S
74	NP_010780.1	PAQR-type receptor	T
75	NP_011049.2	casein kinase YCK3	T
76	NP_012876.1	serine/threonine protein kinase ELM1	T
77	NP_012107.1	formin BNR1	TZ
78	NP_009633.3	Slm4p	U
79	NP_011286.1	She10p	U
80	NP_011620.1	Voa1p	U
81	NP_012343.1	Sop4p	U
82	NP_012932.1	Meh1p	U
83	NP_013026.2	Skg1p	U
84	NP_013220.1	ESCRT-I subunit protein SRN2	U
85	NP_013512.1	Bls1p	U