Universidade de Aveiro Departamento de Ciências Médicas
2021

Catarina Almeida e Silva

Diversidade genética e perfil de suscetibilidade a antifúngicos de dez isolados clínicos de *Candida* spp.

Genetic diversity and antifungal susceptibility profile of ten clinical isolates of *Candida* spp.

Universidade de Aveiro Departamento de Ciências Médicas
2021

Catarina Almeida e Silva

Diversidade genética e perfil de suscetibilidade a antifúngicos de dez isolados clínicos de *Candida* spp.

Genetic diversity and antifungal susceptibility profile of ten clinical isolates of *Candida* spp.

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biomedicina Molecular, realizada sob a orientação científica do Doutor Manuel António da Silva Santos, Professor Associado com Agregação do Departamento de Ciências Médicas da Universidade de Aveiro, e coorientação científica da Doutora Gabriela Maria Ferreira Ribeiro de Moura, Professora Auxiliar do Departamento de Ciências Médicas da Universidade de Aveiro

This work was supported by FEDER (Fundo Europeu de Desenvolvimento Regional) funds through the COMPETE 2020, Operational Programme for Competitiveness and Internationalization (POCI), and by Portuguese national funds via Fundação para a Ciência e Tecnologia, I.P. (FCT) under the project PTDC/BIA-MIC/31849/2017. The iBiMED is supported by FCT funds under UIDP/04501/2020.

o júri

presidente

Professora Doutora Luisa Alejandra Helguero
Professora Auxiliar do Departamento de Ciências Médicas da Universidade de Aveiro

Professor Doutor Manuel António da Silva Santos
Professor Associado C/ Agregação do Departamento de Ciências Médicas da Universidade de Aveiro

Professora Doutora Ana Paula Fernandes Monteiro Sampaio Carvalho
Professora Auxiliar do Departamento de Biologia da Universidade do Minho

agradecimentos

Ao Professor Doutor Manuel Santos e à Professora Doutora Gabriela Moura, pela atenção, disponibilidade e voto de confiança. Obrigada!
À incrível Maria João, pelo seu grande incentivo e papel fundamental na realização deste trabalho, pela sua amizade e carinho. Obrigada por todo o tempo, apoio, dedicação e formação que me proporcionaste.
À Doutora Ana Rita Bezerra e à Rita Guimarães, pela colaboração, apoio, disponibilidade e incentivo. Obrigada!
Às meninas do laboratório, Ana, Carla, e Inês, por me terem acolhido tão bem e pelos conselhos dados. Obrigada!
À minha amiga de sempre e para sempre, Excelentíssima Senhora Dra. Advogada Beatriz Calção, obrigada por toda a motivação, conselhos e desabafos. Seguimos juntas!
À minha amiga Beatriz Melo, sportinguista fiel e sempre pronta a dar tudo de si por quem mais gosta. Obrigada por todo o teu apoio e força!
À minha amiga de infância, que está sempre presente, seja da Polónia, Itália, USA, Dinamarca… Obrigada pela companhia, por me ouvires falar das "bichas" e por me fazeres rir.
Ao Diogo e à Jucalu, são os melhores cunhados/irmãos que podia ter! Obrigada por me fazerem sonhar com os pés assentes na terra, e acreditarem em mim.
Aos meus pais, que mesmo a quilómetros de distância estão sempre lá para mim, a dar-me força, alento e amor. Sem vocês nada disto seria possível, obrigada!
Aos meus avós, o vosso amor e mimo dão-me sempre muita força. Obrigada!
Ao Marco, pela compreensão, incentivo, companheirismo, alegria e amor. Por acreditares comigo e dares-me força para seguir os meus sonhos. Obrigada!

palavras-chave

Candida spp., Sequenciação do genoma total, Polimorfismos, Teste de suscetibilidade a antifúngicos

resumo

Os organismos do género *Candida* são os agentes de infeção fúngica mais comuns, sendo responsáveis por infeções superficiais e sistémicas. Nos últimos anos, verificou-se um aumento de estirpes de *Candida* resistentes a antifúngicos, o que complica o tratamento destas infeções. Com o objetivo de caracterizar 10 isolados clínicos de *Candida*, com base na sua sequência-tipo, polimorfismos de nucleotídeo único, e resistência a antifúngicos, obteve-se a sequência do genoma dos isolados com o MinION (Oxford Nanopore Technologies, UK), e o perfil de suscetibilidade a antifúngicos através de técnicas de microdiluição em caldo e Etest (bioMérieux, França). A análise do genoma dos isolados permitiu perceber que um dos isolados não correspondia a *C. albicans*, sendo *C. glabrata*. Foram identificados SNPs em genes que eram exclusivos para amostras provenientes de exsudatos vaginais, bem como para amostras com origem na cavidade oral. A frequência de genes com SNPs não-sinónimos (missense) envolvidos em processos moleculares e funções biológicas era significativamente superior à frequência base para os mesmos conjuntos de ontologia genética. Todos os isolados eram suscetíveis a todos os antifúngicos testados, e não possuíam nenhum dos polimorfismos associados a resistência descritos na literatura, havendo assim concordância entre o genótipo e o fenótipo.

keywords

Candida spp., Whole-genome sequencing, Polymorphisms, Antifungal susceptibility testing

abstract

Candida species are the most common agents of fungal infection, being responsible for both superficial and systemic infections. In recent years, there has been an increase in antifungal-resistant *Candida* strains, which complicates the treatment of these infections. In order to characterize 10 clinical isolates of *Candida*, based on their sequence-type, single nucleotide polymorphisms, and antifungal resistance, the genome sequence of the isolates was obtained with MinION (Oxford Nanopore Technologies, UK), and the antifungal susceptibility profile was obtained through broth microdilution techniques and Etest (bioMérieux, France). The analysis of the genome of the isolates showed that one of the isolates did not correspond to *C. albicans*, being *C. glabrata*. SNPs were identified in genes that were exclusive for vaginal exudate samples, as well for samples originated from the oral cavity. The frequency of genes with missense SNPs involved in molecular processes and biological functions was significantly enriched to the background frequency in those gene ontology sets. All isolates were susceptible to all antifungals tested, and had none of the resistance-associated polymorphisms found in literature, so there was an agreement between genotype and phenotype.

INDEX

INDEX OF FIGURES

INDEX OF TABLES

INDEX OF APPENDIX

ABBREVIATIONS

ABC	ATP-binding cassette
AFST	Antifungal Susceptibility Test(ing)
ALS	Agglutinin-like sequence
AP	Amphotericin B
AND	Anidulafungin
ATP	Adenosine triphosphate
ATU	Area of Technical Uncertainty
Bcr1	Biofilm and cell wall regulator 1
C. albicans	*Candida albicans*
CBMA	Centro de Biologia Molecular e Ambiental
Cch1	Calcium channel 1
CDR	*Candida* drug resistance
Chr5	Chromosome 5
CLSI	Clinical and Laboratory Standards Institute
CS	Caspofungin
DNA	Deoxyribonucleic acid
DST	Diploid Sequence Type
Ecel1	Extent of cell elongation gene 1
Efg1	Enhanced filamentous growth protein 1
ERG11	Ergosterol biosynthesis gene 11
EUCAST	European Committee on Antimicrobial Susceptibility Testing
Fig1	Mating factor induced gene 1
FL	Fluconazole
g (mg, µg, ng)	Gram (milligram, microgram, nanogram)
GOF	Gain of function
GPI	Glycosylphosphatidylinositol
GWAS	Genome wide association studies
HSP90	Heat shock protein 90
Hwp1	Hyphal wall protein 1
Hyr1	Hyphal regulated cell wall protein 1
ITS	Internal transcribed spacers

IUD	Intrauterine device
L (mL, µL)	Liter (mililiter, microliter)
LOF	Loss of function
LOH	Loss of heterozygosity
M (mM, µM)	Molar (millimolar, micromolar)
MDR	Multidrug resistant
MFS	Major facilitator superfamily
MIC	Minimal Inhibitory Concentration
MLST	Multi Locus Sequence Typing
Mrr1	Multidrug resistance regulator 1
MTL	Mating type locus
MYC	Micafungin
NGS	Next generation sequencing
OD	Optical density
OTU	Operational taxonomic unit
PCR	Polymerase chain reaction
rRNA	Ribosomal ribonucleic acid
RVVC	Recurrent vulvovaginal candidiasis
Saps	Secreted aspartyl proteinases
SNP	Single nucleotide polymorphism
sp. and spp.	Species and “several species”
STAT	Signal transducer and activator of transcription
TAC1	Transcriptional activator 1
Tec1	Transposon enhancement control 1
UPC2	Uptake control protein 2
VVC	Vulvovaginal candidiasis

Other abbreviations will be explained when used in the text.

I. INTRODUCTION

Being the most common causes of fungal infection, *Candida* species are responsible for both superficial and systemic infections. Nearly 90% of these infections are caused by five species: *C. albicans, C. glabrata, C tropicalis, C. parapsilosis,* and *C. krusei*. Three of them, *C. albicans, C. tropicalis* and *C. parapsilosis*, belong to the CTG codon clade, a clade characterized by the CTG codon being translated as serine and not leucine. *C. albicans* is still the most commonly isolated species, but its incidence is decreasing in comparison to other species. *C. glabrata* and *C. krusei* are the two species outside the CTG clade most commonly responsible for infection. *C. glabrata* is more closely related to *Saccharomyces cerevisiae* than to *C. albicans*[1,2]. *C. albicans* and *C. glabrata* are the most commonly isolated among *Candida* species, being responsible, together, for nearly 65% to 75% of all systemic candidiasis[2,3].

An emerging fungal pathogen that has clinical relevance is *Candida auris*. *C. auris* infections are difficult to treat due to high drug resistance, leading to high mortality rates. Phylogenetic studies demonstrate that *C. auris* is distantly related to *C. albicans* and *C. glabrata*. Usually, the clinical isolates show multidrug resistance, characterized by resistance to fluconazole and other azoles, amphotericin B, and echinocandins, hence the concern about the emergence of this fungal pathogen[4].

The misidentification of clinically important *Candida* spp. is a recognised clinical issue. Being important agents of human infection, and with increasing rates of *C. albicans*, *C. glabrata*, *C. krusei*, among others, in local and systemic fungal infections, rapid and correct identification of species plays an important role in the management of candidiasis, as it influences the selection of appropriate antifungal therapy[5,6].

1. *Candida albicans*

Taxonomically, *Candida albicans* belongs to the phylum Ascomycota, class Saccharomycetes. *C. albicans* has an oval yeast-like form and has the ability to grow in three different morphological forms, including budding yeast cells and filamentous forms, being either pseudohyphae or true hyphae[7] (Figure 1). The cells are of two types, “white” and “opaque”; the smooth white colony is composed mainly by budding yeast-like cells, while opaque cells are of an elongated shape and greyish[7,8].

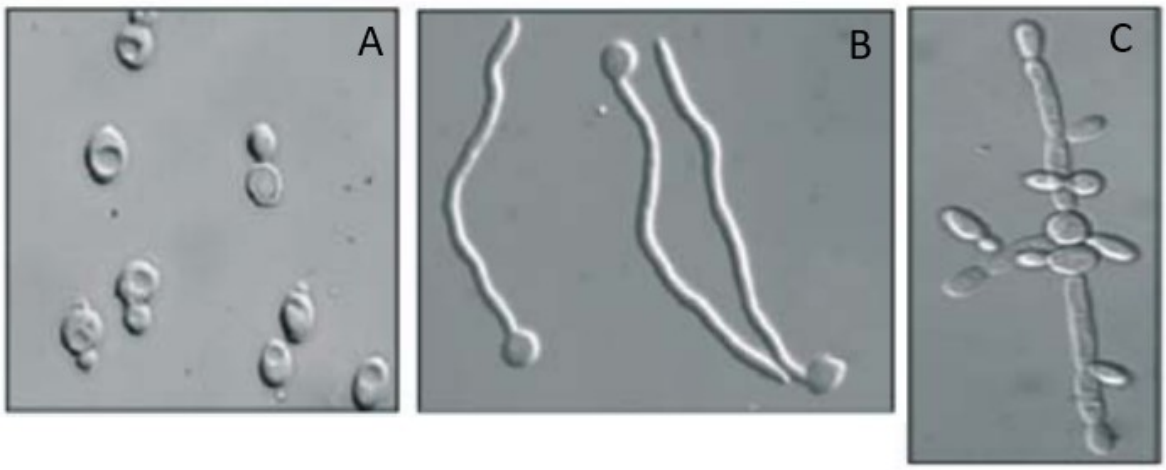


Figure 1. *C. albicans* morphologies. A – yeast; B – hyphae; C – pseudohyphae. Adapted from Kim and Sudbery (2011).

C. albicans is generally a diploid organism (2n) with 8 chromosomes. Its genome size is 13.3-13.4 Mb encoding around 6100-6200 genes[9], but there is great genetic diversity among isolates[10]. *C. albicans* genome has the information that allows it to thrive in the human host while competing with the immune system and other microbiota. It has rarely been isolated in nature away from a host, suggesting that it probably co-evolved with humans and other animals as a commensal organism for millions of years. It has numerous large gene families, some of them encoding known virulence factors, like secreted aspartyl proteinase genes, secreted lipase genes, and agglutinin (*als*) genes. The complete diploid genome assembly for *C. albicans* SC5314, used in research as the reference genome, permitted to estimate a frequency of heterozygosities of approximately 4 polymorphisms per kb, being unevenly distributed across the genome, but having the highest prevalence on chromosomes 5 and 6. The mating-type locus and a region on chromosome 6 that encodes many genes in the *als* gene family are some of the highly polymorphic loci. Additionally, almost half of *C. albicans* genes contain allelic differences, with two-thirds of those polymorphisms being predicted to alter the protein sequence[11].

C. albicans has a parasexual life cycle, in which mating is followed by concerted chromosome loss (Figure 2). The mating-type locus (*mtl*), located on chromosome 5[12], determines the fungus mating type. The *mtl* has two idiomorphs, **a**, that encodes the regulator a1 and a2, and **α**, that encodes the regulators α1 and α2. Both *mtl*a and *mtl*α contain alleles of three additional genes, *pap*, *obp* and *pik28*[13], two of them playing roles in growth and in cellular processes not related to reproduction[14]. The transcription factor a1 and α2 control sexual mating by inhibiting the phenotypic switch that is required in conjugation[14]. Diploid cells (2n) of *C. albicans* are typically heterozygous at *mtl*, however cells can loose their heterozygosity by loss of one copy of chromosome 5, eliminating

the **a** or **α** allele, creating **α** or **a** diploid strain, respectively. The homozygous cells can undergo an epigenetic switch from the white to the opaque phase[15]. Concerning mating, the efficiency is greater in opaque cells than in white cells. This can be explained by the fact that pheromone response is cell-type specific: only opaque a-type cells respond well to pheromone stimulus given that the pheromones produced by α-cells induce projections in the a-cells towards the α-cells. These projections result in the formation of conjugation tubes, allowing for cell and nuclear fusion to form a tetraploid (4n) **a/a/α/α** cell. The tetraploid cells are less stable, so they return to a diploid state by chromosome loss[14], completing the parasexual cycle with no recognized meiosis[15]. In heterothallic mating, this process results in recombinant cells with the potential to have fitness advantages and be drug resistant[14].

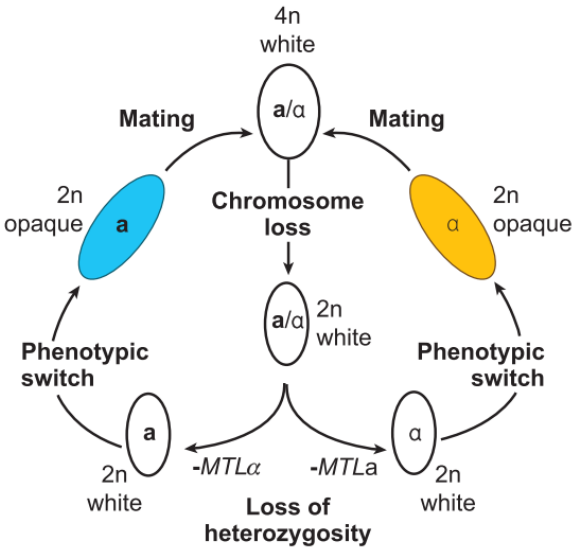


Figure 2. Parasexual cycle of *C. albicans*. Adapted from Noble and Johnson (2007).

Candida albicans is a dimorphic fungus that resides asymptomatically in the gastrointestinal tract of warm-blooded healthy animals, including humans[16]. It is also one of the most common of the opportunistic fungal pathogens. *Candida* infections can be superficial, like oral and vaginal thrush, and chronic mucocutaneous candidiasis, or deep-seated, like *Candida* derived myocarditis and acute disseminated *Candida* septicaemia[16].

2. *Candida glabrata*

While *C. albicans* is a diploid polymorphic fungus, *C. glabrata* is strictly haploid and usually grows in the yeast form only[2].

Despite frequent high levels of intrinsic and acquired resistance to the azoles class of antifungals, *C. glabrata* does not have the same virulence mechanisms present in *C. albicans*[3]. In fact,

due to the absence of an invasive growth form, it does not cause extensive epithelial damage[2,17]. The fungus can survive and replicate inside macrophages until they burst; this has been hypothesized to be a way of tissue invasion, however its mechanisms of invasion are not well known[3]. Since *C. glabrata* is able to suppress, or is unable to stimulate neutrophil attraction, it does not incite a strong immune response by its host[2].

The *C. glabrata* genome size is in the order of 12 Mb, displaying 12 chromosomes and one mitochondrial genome; even though only approximately 6% of *C. glabrata* genome is annotated on Candida Genome Database (CGD)[18], it contains all the genes required for mating and meiosis[1]. Whole-genome sequence comparison of diverse lineages that undergo recombination failed to clarify if recombination occurs in the course of human infection or in commensalism. The genome also contains several members of a gene family encoding epithelial adhesin proteins (*epa* genes), considered to be a key factor in the yeast pathogenicity toward humans. SNPs analysis shows an enrichment in genes of cell wall proteins and high levels of genetic variation, although because it is a haploid organism, loss of heterozygosity does not have a role in this genetic variation; aneuploidies can be considered[19].

3. Epidemiology

C. albicans is the most common etiologic agent of human fungal disease. It is estimated that 25% to 50% of healthy humans carry *Candida* as part of the normal microbiota of the mouth, with *C. albicans* accounting for 70% to 80% of isolates. The predominant source of infection caused by *C. albicans*, from superficial mucosal and cutaneous disease to hematogenous dissemination, is the patient; most types of candidiasis represent endogenous infection in which the normally commensal host microbiota takes advantage of the "opportunity" to cause infection. Among the various species of *Candida* capable of causing human infection, *C. albicans* predominates in most types of infection. Infections of genital, cutaneous, and oral sites almost always involve *C. albicans*[7].

3.1. Vulvovaginal candidiasis

Vulvovaginal candidiasis is the second most common cause of vaginitis after bacterial vaginosis, and it is estimated that 10% to 15% of asymptomatic women are colonized by *Candida* species, 70% to 75% of women will have an episode of vulvovaginal candidiasis (VVC) during their lifetime, 50% of these women will experience a second VVC episode, and 5% to 10% of all women will develop recurrent vulvovaginal candidiasis (RVVC)[20]. Epidemiologic studies show that women in reproductive age have higher incidence of VVC than women at menopause[20].

3.2. Oral candidiasis

C. albicans colonises the oral cavity of up to 75% of the population; this colonization is normally benign in healthy individuals; however, immunocompromised individuals can suffer from persistent infections of the oral cavity. Oral candidiasis can affect the oropharynx and the esophagus of people with some kind of disfunction of the adaptive immune system[21,22]. *C. glabrata* accounts for approximately 5% to 8% of isolates recovered from infected patients[23].

3.3. Systemic candidiasis

Candida species can cause infections of any organ system, ranging from superficial mucosal and cutaneous candidiasis to widespread hematogenous dissemination targeting organs such as the liver, spleen, kidney, heart, and brain[7]. Systemic candidiasis is associated with high mortality rate, even with first line antifungal therapy. Risk factors include neutropenia and damage of the gastrointestinal mucosa, and the use of central venous catheters, which allow direct access of the organism to the bloodstream, the application of broad-spectrum antibacterials, which enable fungal overgrow, and trauma or gastrointestinal surgery, given that it disrupts mucosal barriers[21].

4. Pathogenicity

Candida species' pathogenicity is mediated by virulence factors, such as adhesion, biofilm formation, contact sensing and thigmotropism, extracellular hydrolytic enzyme production, polymorphism and phenotypic switching.

4.1. Adhesion

The primary event is the adhesion of *Candida* species to epithelial cells, mediated by cell surface physicochemical properties and promoted by specific cell surface proteins – adhesins. In *C. albicans*, adhesins are encoded by the agglutinin-like sequence (*als*) gene family, consisting of eight members (Als1p-Als7p and Als9p) that are glycosylphosphatidylinositol (GPI)-linked to the β-1,6-glucans of the cell wall[24]. These proteins recognize host ligands, like serum proteins and components of host tissues, or promote the binding to abiotic surfaces (medical devices, such as intrauterine device (IUD)) through hydrophobic interactions[20]. Although expression of *als* genes differs according to the fungal morphology and body site, Als3p appears to be key in the adherence of *C. albicans* hyphae to epithelial cells and the subsequent invasion of host cells; *als3* is upregulated during infection of oral and vaginal epithelial cells, while blocking of Als3p or preventing *als3* expression causes a massive reduction in epithelial adhesion[24].

C. glabrata has a large protein family of adhesins, the Epa proteins that allow the cell's attachment to epithelial cells and macrophages[3]. The expression of Epa proteins is regulated by Silent information regulator(Sir)- complex mediated transcriptional silencing, which is thought to be the mechanism influencing the sensing of host's environment[2]. The most important adhesins are Epa1, Epa6, and Epa7; deletion of *epa1* gene decreases *C. glabrata* adhesion to host epithelial cells in vitro[25].

C. albicans and *C. glabrata* have evolved different and specific adhesins that count on different indications to detect the presence of host's cells; once it happens, adhesins are expressed to link to host cells[2].

4.2. Biofilm

After adhesion is established, biofilm formation can start. This is an important virulence factor for *Candida* species since it confers phenotypic characteristics, including resistance to antifungal agents, host defense mechanisms and physical and chemical stress. Also, biofilm cells show metabolic cooperation, community-based regulation of gene expression and the ability to withstand competitive pressure from other organisms, contributing to their survival in hostile environment conditions[20]. The most common substrates include catheters, dentures, and mucosal cell surfaces[21]. *Candida* species can form biofilms on vaginal epithelium and also on IUDs, promoting VVC[20]. The resistance of biofilms to the conventional antifungal therapy may prevent the complete eradication of these microorganisms from the vaginal and vulvar area, and might explain the recurrence of VVC[20]. Biofilms come into form by a sequential process that includes the adherence of yeast cells to the substrate, their proliferation, hyphal cell formation in the upper part of the biofilm, accumulation of extracellular matrix material and the dispersion of yeast cells from the biofilm complex[21]. Some transcription factors have been associated to the control of biofilm formation, such as Bcr1, Tec1 and Efg1; also the heat shock protein Hsp90 has been identified as a key regulator of dispersion of *C. albicans* biofilms and as a requirement for biofilm antifungal drug resistance[21].

C. glabrata clinical isolates have the ability to form compact biofilms, structured in different multilayers with proteins, β-1,3 glucans, and ergosterol in their matrixes, providing them with high tolerance to antifungal treatments[25,26]. Biofilm formation for *C. glabrata* starts with attachment of yeast cells followed by cell division, which leads to forming the basal layer of anchoring microcolonies, with subsequent biofilm maturation. The genes so far described to be involved in biofilm formation are *sir4*, telomere-binding (*rif1*), *epa6*, and serine-threonine protein kinase gene (*yak1*)[25].

4.3. Contact sensing and thigmotropism

Contact sensing is an important factor that triggers hypha and biofilm formation in *C. albicans*; upon contact with a surface, yeast cells switch to hyphal growth, and can invade into the mucosal barrier. The contact to the surfaces also induces biofilm formation. The hyphal growth can be directional – thigmotropism – if the surfaces have particular topologies, like ridges. *C. albicans* thigmotropism is required for damage of epithelial cells and normal virulence in mice, and is regulated by extracellular calcium uptake through the calcium channels Cch1, Mid1 and Fig1, and also an additional mechanism, the polarisome Rsr1/Bud1-GTPase module[21,27]. The ability of external cues or environmental factors to trigger the orientation of hyphal growth is therefore relevant for *C. albicans* capacity to infiltrate between human cells during tissue invasion[27].

4.4. Secreted hydrolases

C. albicans secrete hydrolytic enzymes with an important role in adhesion, tissue penetration, invasion and destruction of host tissues[20]. Secreted hydrolases are also thought to enhance the efficiency of extracellular nutrient acquisition[21]. These enzymes are secreted aspartyl proteinases (Saps), phospholipases, lipases and haemolysins. Saps facilitate adhesion to host tissues and its damage is intrinsically related to modifications in the host immune response. Phospholipases hydrolyse ester bonds in glycerophospholipids contributes to host-cell membrane damage and to adhesion of yeasts to host tissues. Lipases are involved in the hydrolysis of triacylglycerols and its activity has been associated to *C. albicans* adhesion, damaging of host tissues and negative affect on immune cells. Haemolysins degrade haemoglobin, facilitating the iron recovery, which is essential to survival and persistence in the host[20].

C. glabrata shows inability to secret proteases, however it has the capacity to secrete phospholipases, lipases and haemolysins, that just like for *C. albicans*, contributes towards pathogenicity[25].

4.5. Polymorphism and phenotypic switching

The reversible morphological alternation between unicellular yeast cells and the filamentous phase is an important virulence factor, since filamentous forms have more mechanical strength, enhancing colonization and invasion of host tissues and demonstrate increased resistance to phagocytosis[20]. *C. albicans* colonies can switch among different phenotypes at high frequency. The phenotypic switching changes several virulence characteristics, including hyphae formation, drug

resistance, adhesion and Saps secretion, which can affect survival in specific anatomical sites and promoting infection. The transition between yeast and hyphal growth forms is named dimorphism, and it has been proposed that both forms are important for pathogenicity. Environmental characteristics, such as pH, affect *C. albicans* morphology; at low pH (<6) *C. albicans* tend to grow in the yeast form, while at higher values (>7) hyphal grow is induced. Also, starvation, the presence of serum or N-acetylglucosamine, physiological temperature and CO_2 promote the formation of hyphae. Quorum sensing, a mechanism of microbial communication, has been shown to regulate morphogenesis; due to this, high cell densities (>10^7 cells/ml) promote yeast growth, and low cell densities (<10^7 cells/ml) promote hyphal formation. Interestingly, hyphal formation is associated to the expression of some genes encoding virulence factors that are not directly involved in hyphal formation per se; hypha-linked proteins include the hyphal wall protein Hwp1, the agglutinin-like sequence protein Als3, the secreted aspartic proteases Sap4, Sap5 and Sap6, and the hypha-associated proteins Ecel1 and Hyr1[21].

C. glabrata switching represents graded systems; when growing on an indicator agar containing $CuSO_4$, the yeasts reveal a spontaneous, reversible and high-frequency switching system, presenting three different colony phenotypes: white, light brown, and dark brown. This switching regulates the expression of selected genes, like the metallothionein gene *mtii*, which transcription level regulation is selective and in a graded hierarchy[28].

Despite all the virulence factors that mediate *C. albicans* pathogenicity, they are influenced by specific environmental conditions of each anatomical site. That said, the disturbance of the normal environment of the female genital tract, or of other anatomical site, may increase the risk of infection by *C. albicans*[20]. The lack of hypha formation in *C. glabrata* is, most likely, a contributing factor for its low virulence profile, especially considering that this characteristic allows *C. albicans* to have increased adherence and tissue invasion, and also increased proteolytic enzyme activity[23].

5. Antifungal drugs used in human medicine and resistance

5.1. Inhibitors of ergosterol biosynthesis - azoles

Fluconazole is an antifungal that belongs to the largest family of antifungal drugs, the azoles[29]. This family of antifungals works by inhibiting the cytochrome P450 enzyme lanosterol demethylase (14α-demethylase), which converts lanosterol to ergosterol[29–31]. Ergosterol is an important component of fungal cell membranes so the inhibition of ergosterol biosynthesis pathway by fluconazole is toxic for the fungus[31]. Lanosterol demethylase, encoded by the *erg11* gene, is inhibited due to the binding of the free nitrogen atom of the azole ring to an iron atom within the

heme group of the enzyme, blocking the activation of oxygen and in turn the demethylation of lanosterol. The accumulation of methylated sterols in the fungal cellular membrane arrests cellular growth, therefore fluconazole has a fungistatic action. Because of this, the treatment provides room for the fungus to develop acquired resistance to the drug[31].

The molecular mechanisms involved in fluconazole resistance include genes of the ergosterol biosynthesis pathway (drug target overexpression or alteration, bypass pathways), drug transporters (efflux pump overexpression), changes in ploidy and loss of heterozygosity (LOH)[29,31,32]. By overexpressing *erg11*, the gene encoding for the azoles' target, *C. albicans* increase the production of the enzyme lanosterol demethylase and dilute the fluconazole activity, which therefore results in resistance. This upregulation is often related with a gain of function (GOF) mutation in Upc2p, a transcription factor that is induced consequent to ergosterol depletion[31].
Still in the *erg11* gene, point mutations in its coding region results in amino acid substitutions that change the structure of the protein lanosterol demethylase, and consequently contributing to a less effective binding of the azoles[29,31]. These point mutations result in fluconazole resistance by reduced binding affinity.

Alterations in sterol biosynthesis, or in other words, the development of bypass pathways, can occur in order to prevent alterations in the cell membrane and the build-up of toxic products, which results in resistance to fluconazole, since these new pathways are not interrupted by the azoles[29]. Even though this mechanism of resistance is less frequent, it results from loss of function (LOF) mutations in the *erg3* gene. This blocks the accumulation of the toxic sterol intermediates when *erg11* is inhibited by the azoles[33]. Consequently, the enzyme delta 5,6-sterol desaturase is inactivated, allowing the cell to bypass synthesis of methylated sterol products, therefore minimizing fluconazole activity[31].

Efflux pumps in the fungal cell wall allows them to pump out the antimycotic drug, resulting in the failure of the drug to accumulate intracellularly and consequently there is resistance[29,31]. This is possible due to transport/efflux proteins, the ATP-binding cassette (ABC) transporters and the major facilitator superfamily (MFS) transporters. The upregulation of these transporters mediates resistance to fluconazole[31,34]. Several fluconazole resistant *Candida* isolates overexpress the *cdr1* and *cdr2* genes, that encode for ABC proteins, and it has been connected to azoles resistance. The upregulation of these genes is a consequence of GOF mutations in the transcriptional regulator *tac1*[31,32]. There are at least 19 different GOF point mutations confirmed in *tac1* for *C. albicans*[31].

The *mdr1* gene encodes for one specific MFS transport protein, and its overexpression has been linked to fluconazole resistance. The regulation of the expression of this gene is done by the

transcription factor *mrr1*[31,32], and there are reported at least 15 different GOF point mutations within this factor for *C. albicans*[31].

Aneuploidy and loss of heterozygosity (LOH) are also involved in fluconazole resistance in *C. albicans*. It was found that aneuploid chromosomes are seven times more common in fluconazole resistant isolates of *C. albicans* when compared with fluconazole susceptible isolates; more specifically, isochromosome formation of chromosome 5 (chr5) is associated with fluconazole resistance due to the duplication of the segment which grants additional copies of the *erg11* gene, its transcriptional factor *upc2* and also the efflux pump regulator *tac1*[31,35]. LOH is implicated in fluconazole resistance when associated with the genes *erg11*, *tac1* and *mrr1*; the phenomenon starts with point mutations in the heterozygous state that gradually evolves to homozygosity or deletions of one allele, followed or not by recombination[31].

C. glabrata is generally characterized by presenting high levels of both intrinsic and acquired resistance to azoles, as a consequence of multidrug resistance transporters overexpression, activated by the transcription factor Pdr1. Gain of function mutations within *pdr1* gene upregulate the drug efflux pumps *Cdr1*, *Cdr2*, *Snq2* and *Qdr2*, thus conferring most of the acquired resistance in *C. glabrata*; this was also found to have a role in enhancing virulence for this pathogen[3]. Loss of mitochondria also leads to azole resistance by *C. glabrata*, which is correlated to the upregulation of *cdr1* gene[36].

The absence of the adhesin Epa3 was shown to increase *C. glabrata* susceptibility to azoles, probably explained by the fact that Epa3 promotes cellular aggregation, protecting the yeast cells from extracellular drug concentration[3].

The single nucleotide polymorphisms (SNPs) associated to azoles resistance found in the literature are depicted in Table 1.

Table 1. Missense mutations associated to azoles-resistance in *C. albicans* and *C. glabrata*.

Specie	Resistance	Gene	Nucleotide substitution	Amino acid substitution	Literature
C. albicans	Azoles	*erg11*	214T>C	F72L	Favre (1999)[37]
			394T>C	Y132H	Favre (1999)[37]
			1349G>A	G450E	Favre (1999)[37]
			1309G>A	V437I	Favre (1999)[37]
			622G>A	E208K	Spettel (2019)[38]
			1574C>T	T525I	Spettel

					(2019)[38]
			428A>G	K143R	Manastir (2011)[39]; Flowers (2015)[40]
C. albicans	Azoles	*erg3*	503C>T	A353T	Morio (2012)[41]
			986C>T	T329S	Morio (2012)[41]
			571T>C	S191P	Spettel (2019)[38]
			782G>A	G261E	Spettel (2019)[38]
C. albicans	Azoles	*tac1*	2929A>G	N977D	Coste (2006)[42]
			2810C>T	S937L	Spettel (2019)[38]
			2218A>G	N740D	Siikala (2010)[43]
			2939G>A	G980E	Coste (2009)[44]
C. glabrata	Azoles	*pdr1*	3235G>A	G1079R	Ferrari (2009)[45]
			2626G>T	D876Y	Ferrari (2009)[45]
			3236G>T	G1079V	Ferrari (2009)[45]
			1043G>A	G348D	Ferrari (2009)[45]
			871T>C	L291P	Ferrari (2009)[45]
			1042G>A	G348S	Ferrari (2009)[45]
			1114T>C	Y372H	Ferrari (2009)[45]
			1037G>A	G346D	Ferrari (2009)[45]
			862C>G	H288D	Spettel (2019)[38]
			1037G>C	G346A	Ferrari (2009)[45]

5.2. Fungal membrane disruptors – polyenes

Polyenes are organic molecules with an amphiphilic structure, which allows these molecules to bind to the lipid bilayer and form a complex with the ergosterol, producing pores; this promotes

the disruption of the cell membrane, the leakage of the cytoplasmic contents and oxidative damage, which results in fungal cell death. The polyene amphotericin B is active against most yeasts and filamentous fungi, and is therefore recommended for the treatment of infections caused by *Candida albicans*; in particular, amphotericin B binds to ergosterol and forms an extra membranous fungicidal sterol sponge, destabilizing membrane function[34]. Amphotericin B is the most used polyene for the treatment of systemic infections, being administered intravenously; however, polyenes have a certain affinity for cholesterol, which explains the high toxicity and side effects associated to these antifungals[34].

Resistance to polyene is rare, and in most cases is due to a decrease in either the amount of ergosterol or a change in the target lipid; it is frequently linked to changes in both *erg3* and *erg6*. Disruption of these genes causes decreased ergosterol levels, because instead of producing ergosterol, both *C. albicans* and *C. glabrata* produce ergosterol-like compounds that have less affinity for amphotericin B[36,46]. Another mechanism involved in polyene resistance in order to reduce the effective drug concentration is the ability to form biofilms; biofilms restrict the penetration of polyenes by forming a diffusion barrier sequestering the antifungal agent[34].

5.3. Fungal cell wall synthesis – echinocandins

Echinocandins are semisynthetic lipopeptides derived from fungal natural products, and the only available antifungal drugs targeting cell wall: caspofungin, micafungin and anidulafungin. They have fungicidal activity against many strains of *Candida*: both micafungin and anidulafungin are licensed for the treatment of invasive and oesophageal candidiasis[34]. Their action is as non-competitive inhibitors of β-(1,3)-D-glucan synthase enzyme complex, targeting specifically the FsK1 subunit, leading to the disruption of the structure of growing cell walls, which results in osmotic instability and fungal cells death. Echinocandins have good safety profiles and low toxicity, since their target is absent in mammalian cells and their interactions with other drugs are minimal. Also, their use in combination with azoles or polyenes can enhance their action against *Candida* strains[34].

Resistance to echinocandins appear to be due to combinations of mutations in the *fks1* or *fks2* genes and ABC transporters upregulation. These mutations are located in two different hot spot regions of these genes, HS1 and HS2, and were reported in *C. albicans*; the mutations change the kinetics of the glucan synthase, which results in significantly higher inhibitory concentrations and inhibition constant[34]. *fks1* and *fks2* point mutations have been observed in *C. glabrata*, arising spontaneously in the presence of echinocandin selection pressure[36].

The development of novel resistance profiles has been reported, affirming its simultaneous

resistance to at least two different classes of antifungal agents, that is, some strains are multidrug resistant (MDR). In the particular case of *C. albicans*, loss-of-function point mutations or simultaneous mutations in *erg2, erg3, erg5* and *erg11* result in MDR to azoles and polyenes. Also, there has been an increase in observed azole- and echinocandins-resistant *C. albicans* isolates[34]. The SNPs associated to echinocandin resistance found in the literature are depicted in Table 2.

One of the preferred mechanisms for evolutionary advantage regarding antifungal resistance, for both *C. albicans* and *C. glabrata*, appears to be the occurrence of mutations, which then allows for natural selection of a SNP or a set of SNPs that confers advantages under selective pressures[3].

Table 2. Missense mutations associated to echinocandins-resistance in *C. albicans* and *C. glabrata*.

Specie	Resistance	Gene	Nucleotide substitution	Amino acid substitution	Literature
C. albicans	Echinocandins	*gsc1*	1922T>G	F641C	Wiederhold (2011)[47]
			1933T>C	S645P	Garnaud (2015)[48]
			1946C>A	P649H	Dudiuk (2015)[49]
			2086A>G	M696V	Spettel (2019)[38]
C. albicans	Echinocandins	*erg3*	1057G>A	A353T	Morio (2012)[41]
C. glabrata	Echinocandins	*fks2*	1992C>A	P667T	Spreghini (2012)[50]
			1987T>C	S663P	Garnaud (2015)[48]

6. Methods for genome analysis

Genomic studies are usually done using next-generation sequencing platforms, and there are established experimental and bioinformatic protocols for sample processing, nucleic acid isolation, sequence target amplification, library preparation, sequence data processing and statistical analysis[51].

Whole-Genome Sequencing (WGS) allows for the analysis of entire genomes, providing a base-by-base review of the genome and capturing both large and small potential causative variants, that can be used in further follow-up studies of gene expression and regulation mechanisms. Genomes of the same species share a common set of genes, known as the core genome; variability can occur, including point mutations and differences in genome content. WGS can be done by mapping reads or assembled contigs to a reference genome of the species being studied. Although

this type of approach can give very high resolution, the reference genome used must be closely related to the sequenced samples (to the level of species or even strains, in case of highly variable species), in order to reduce the probability of mismapping and increase the reference genome's regions to which reads will be mapped against[52], the so-called coverage.

Nanopore-based sequencing strategies are recently emerging in the field, and involve strand sequencing where intact DNA circulates through the nanopore base by base – Oxford Nanopore Technologies (ONT)[53]. Within the benefits of this technology there is the detection of base modifications, as by performing single-molecule sequencing of DNA and RNA, nanopore technology can detect modifications on individual nucleotides, including cytosine methylation in genomic DNA; real-time targeted sequencing is also a benefit since it can reduce significantly the time needed from biological sampling to data analysis – for example, MinION device is cheap, portable, small-sized and the library preparation is simple, compared to other platforms, conferring an unique advantage: on-site sequencing[54]; another advantage of using ONT is extending read lengths that exceed the ones of dominant NGS platforms; it also assists in the reliable detection of structural variants with only a few hundred reads in opposition with the millions of reads often required when using NGS platforms[53]. ONT can bring advantages for the challenge of fungal genome assembly due to being able to offer ultra-long reads and resolving repeat-rich sequences, which is the case of the fungal genome, since it has a high repeat content[55]. ONT has been tested for genome sequencing research of bacteria, viruses, and yeast, showing a great potential for infectious disease diagnosis[54]. In summary, MinION advantages rely on using a portable, real-time, simple, fast, and low-cost platform that supports long reads. The greatest disadvantage is its higher error rate that it poses a challenge due to higher rates in the base called reads – 10% when comparing to 1% for Illumina), and originates a significant number of insertions and deletions, that are difficult to correct[56,57].

Multi Locus Sequence Typing (MLST) is a method used to differentiate and characterize microbial isolates for epidemiological purposes[58], using the sequences of internal fragments of, normally, seven house-keeping genes. The different sequences present within a species are assigned as distinct alleles for each house-keeping gene, and for each isolate, the alleles at each of the seven loci define the allelic profile or sequence type (ST)[59].

If two isolates have different STs, then the strains are different; if two strains have the same ST, then the strains are not distinguishable by MLST, however they are not necessarily identical strains, since this methodology gives no information about the global extent of SNP differences between the two sequences[60].

MLST scheme databases for *C. albicans* (*aat1a, acc1, adp1, mpib, sya1, vps13*, and

zwf1b[58,60]), and *C. glabrata* (*fks, leu2, nmt1, trp1, ugp1* and *ura3*[61]), are hosted in PubMLST (Public databases for molecular typing and microbial genome diversity) which is publicly available (https://pubmlst.org/)[62].

The five largest clades assigned by MLST for *C. albicans* are clade 1, clade 2, clade 3, clade 4, and clade 11, considered to be the most consistent clades on the MLST database. Certain STs occur with higher frequency than others within the MLST clades[63]. Molecular strain typing allows to demonstrate that certain types are more commonly associated with invasive disease than other types. There is a significantly greater proportion of isolates associated with superficial and mucosal infections and commensal carriage in clade 1 against other clades. They are found more frequently and have a more general distribution than other types since they are better adapted to cause human disease[60,63]. Flucytosine resistant isolates harbouring a mutation in the *fur1* gene are very common in clade 1[60]; the same mutation has not been found in isolates resistant to flucytosine from other clades[63]. Similarly, there is a considerable number of isolates resistant to amphotericin B in clade 4[60].

Regarding *C. glabrata*, 7 major groups have been identified[64]. There is a significant bias in the geographical distribution of the groups: in groups 1 and 3, European strains are more common, in group 3 strains from the USA are over-represented, Japanese strains are more common in group 5, and strains originated from South America are over-represented in group 6 and 7[64].

Gene Ontology (GO) represents the current state of knowledge within the biological domain, being the world's most comprehensive source of information about the function of gene and gene products, with respect to three aspects; (i) molecular function; (ii) cellular component; and (iii) biological process. Ontologies consist in a set of terms with relations that operate between them, that are composed by a definition, a label, a unique identifier, and other elements[65,66]. The utility of GO annotation goes for making predictions of functions for gene products across numerous species; GO enrichment analysis provides a way for identifying processes that are statistically overrepresented among a gene set of annotations[67]. At the CGD[18], GO enrichment analysis provides a powerful means for evaluating function and localization terms, and annotations within the biological process branch[67].

7. Antifungal Susceptibility Testing (AFST)

Antifungal susceptibility testing (AFST) is a tool in clinical microbiology to detect antifungal resistance and to facilitate the choice of the optimal antifungal agent for a fungal infection[68]. In addition to guide patient therapy, AFST is also an important tool to inform epidemiological studies on antifungal resistance, and to study the *in vitro* activity of new antifungal agents[69]. AFST allows the

determination of the minimum inhibitory concentration (MIC), which is the lowest drug concentration that inhibits the growth of organisms within a defined period[69].

7.1. Methods to perform AFST

The reference techniques of antifungal susceptibility testing are divided in dilution methods and agar-based methods[70]. The European Committee on Antimicrobial Susceptibility Testing (EUCAST) and the Clinical Laboratory Standards Institute (CLSI) have standardized methodologies in order to perform antifungal susceptibility testing[69,71].

Broth microdilution methodology measures the ability of the microorganism to produce sufficient growth in microdilution plate wells of broth culture media covering serial dilutions of the antimicrobial agent[69].

The Etest®, by bioMérieux, is a commercially available product for antifungal susceptibility testing, that consists on thin strips of plastic that contains a predefined gradient of antifungal drug on one side, while the other side is marked with the concentration scale. The strip is placed on a confluent lawn of cells and then incubated. During incubation, the antifungal diffuses into the agar. The MIC value is read by interpreting the zone of inhibition – the point at which the ellipse-shaped growth of cells intercepts with the strip indicates the MIC value[68].

Other methods that can be used for antifungal susceptibility testing are disk diffusion testing[68], and commercially available products like Sensititre YeastOne® panel (ThermoFisher Scientific, Waltham, MA)[70].

8. Objectives

The aim of this dissertation was to study the genetic diversity of ten clinical strains previously identified as *Candida albicans* isolates. For this, the specific objectives were, to:

1 - Establish the antifungal susceptibility profile of the isolates against distinct classes of antifungal drugs used in clinical therapeutics.

2 – Sequence the genome of the isolates using next generation sequencing methods.

3 - Determine the isolates' genetic profiles based on MLST.

4 - Identify single nucleotide polymorphisms in the isolate's genome.

5 - Identify previously described single nucleotide polymorphisms (SNPs) associated to antifungal drugs resistance.

6 - Categorise the GoTerms associated with the SNPs identified, and link this to predicted mechanisms of resistance.

II. METHODOLOGY

1. Strains and growth conditions for DNA extraction and sequencing

Candida spp. clinical isolates were kindly provided by Departamento de Biologia da Universidade do Minho and by the Centro de Biologia Molecular e Ambiental (CBMA). All *Candida* sp. used in the study are listed in Table 3.

The pre-requisite for using these strains was that they were clinical isolates; the isolates were identified to the species level by Professor Célia Pais's group by growing the yeasts in CHROMagar™ Candida, followed by PCR-based methodologies. We did not know if the isolates provided were virulent (isolated from infections) or exist as commensal organisms at the time of collection.

Candida sp. isolates were grown at 30°C in YPD (Formedium) solid medium with 2% agar, for 96 hours. Strains were pre-inoculated in YPD broth at 30°C, overnight. Optical density was measured and 100 ml of YPD broth were inoculated with cell cultures with OD_{600} at 0.02 (measurement performed in the spectrophotometer Microplate Manager version 6.3) and grown overnight until reaching the log growth phase.

Table 3. *C. albicans* strains used in the study.

Strains	Human source	Reference number in the study
SER20	Haemoculture	1
SER7	Haemoculture	2
S142	Oral cavity	3
S145	Oral cavity	4
S005a	Oral cavity	5
5718fs	Vaginal exudate	6
81	Haemoculture	7
7333fs	Vaginal exudate	8
S038	Oral cavity	9
7248fs	Vaginal exudate	10

2. Genomic DNA extraction

For DNA extraction, the QIAGEN Genomic DNA Purification kit with Genomic-Tips 100/G was used (Qiagen, Germany) with alterations in incubation and centrifugation times, described below.

Yeast cells were harvested from a total of 100 ml of culture by centrifugation. Cells were resuspended in 4 ml of TE buffer (10 mM Tris•Cl, pH 8.0; 1 mM EDTA, pH 8.0, per liter of solution) by vortexing, and cells were pelleted by centrifugation. The supernatant was discarded and the pellet resuspended in 4 ml of Buffer Y1 (Qiagen, Germany) and vortexed. Subsequently, 250 µl of lyticase (1000U/ml) were added and the suspension was incubated for 2 hours at 30°C. After incubation, the

mixture was centrifuged for 20 minutes. The supernatant was discarded and cells were resuspended and mixed in 5 ml of Buffer G2 (Qiagen, Germany) with RNase A (200 µg/ml). Afterwards, 97 µl of QIAGEN Proteinase K stock solution was added to the cells and mixed by inverting the tube and vortexing. The mixture was incubated for 2 hours at 50°C. After incubation, the suspension was centrifuged for 20 minutes. The supernatant was retained, and the pellet was discarded.

The Genomic-Tips 100/G were equilibrated in 4 ml of Buffer QBT (Qiagen, Germany) and were allowed to empty by gravity flow. Samples were vortexed for 10 seconds at maximum speed and applied to the equilibrated QIAGEN Genomic-tip, where they entered the resin by gravity-flow; flow was left unassisted until tip emptied. The QIAGEN Genomic-tip was washed twice with 7.5 ml of Buffer QC (Qiagen, Germany), and flow was left unassisted until tip was cleared. The QIAGEN Genomic-tip was then placed into a new Falcon tube, where the genomic DNA was eluted with 5 ml of prewarmed Buffer QF at 50°C (Qiagen, Germany) and the tip was allowed to drain by gravity flow. The QIAGEN Genomic-tip was then removed, and 3.5 ml of isopropanol (0.7 volumes, at room-temperature) was added to the eluted DNA. The DNA was precipitated by gently inverting the tube 10 to 20 times, and spooled using a sterile loop. The spooled DNA was immediately transferred to a microcentrifuge tube containing 100 µl of sterile Mili-Q water. The tube was left open overnight, at 4°C, allowing for the evaporation of the isopropanol. Samples were stored at 4°C until further use.

3. Whole-genome sequencing using Oxford Nanopore Technologies

Whole genome sequencing was performed twice to improve yield of sequencing reads.

3.1. DNA quantification

Genomic DNA and DNA libraries quantification was performed using Qubit® 2.0 Fluorometer (Invitrogen by Life Technologies, USA), following Qubit® dsDNA BR Assay (http://tools.thermofisher.com/content/sfs/manuals/Qubit_dsDNA_BR_Assay_UG.pdf).

3.2. Genomic library preparation

Genomic libraries were prepared using the Ligation kit SQK-LSK109 (ONT, UK) and the native barcoding kit EXP-NBD104 (ONT, UK), to allow for multiplexing of samples in the same sequencing run. This was done according to ONT protocol Native barcoding genomic DNA (with EXP-NBD104 and SQK-LSK109) (available at https://store.nanoporetech.com/media/wysiwyg/pdfs/EXP-NBD104/native-barcoding-genomic-dna-NBE_9065_v109_revJ_23May2018.pdf), with few modifications: (i) the input DNA concentration (40.8 ng/µl) was doubled to increase data generation;

(ii) incubation times with reaction mixtures and Agencourt AMPure XP beads (Sigma-Aldrich, Germany) were doubled to ensure maximum recovery of DNA; (iii) incubation with elution buffer was done at 37°C for 10 minutes to improve the recovery of long fragments.

3.3. Sequencing run with ONT MinION

The sequencing run was performed twice for each sample using the run options default parameters with live basecalling through Guppy (fast model), demultiplex option on, and standard 48-hour run script (Figure 3). Base called reads were classified either pass or fail depending on their modal per base quality score, which was set to a minimum Q score of 7.

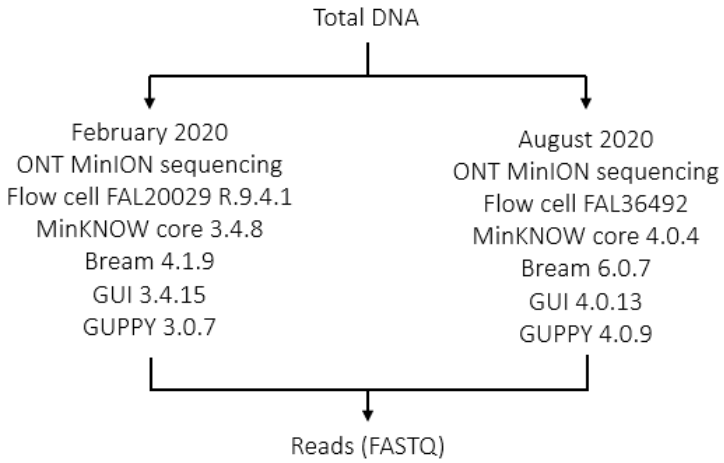


Figure 3. Sequencing run characteristics.

4. Bioinformatics analysis

Quality and metrics of the sequencing runs were verified using Nanoplot (v1.27.0)[72]. To improve the accuracy of bases in reads, the Canu (v2.0)[73] correction option with default parameters was used to correct raw reads.

de-novo assembly of Canu corrected reads in contigs was performed using Flye (v2.7)[74]. Polishing consisted of one round of Racon (v1.4.3[75]; for an initial correction in the raw contigs), followed by one round of Medaka (v1.2.1[76]; to obtain a consensus sequence) and four iteractions of Pilon (v1.23[77]; for final polish). The genome assemblies obtained were used for MLST analysis (Figure 4, left).

Using Minimap2 (v2.17)[78], Canu corrected reads were mapped against the reference genomic sequences of strains *Candida albicans*, SC5314 haplotype A (version A22-s07-m01-r100) or *C. glabrata* CBS138 (version s03-m01-r04), both available at the *Candida* Genome Database (http://candidagenome.org/). Sequencing depth for each sample was calculated.

To look for similar genomic sequences to that of sample 02 (SER7) and determine its species identity, the blastn program within the BLAST+ software (v2.10.1)[79] was used. The 10 best sequence matches with E values lower than 10e-5 were retrieved.

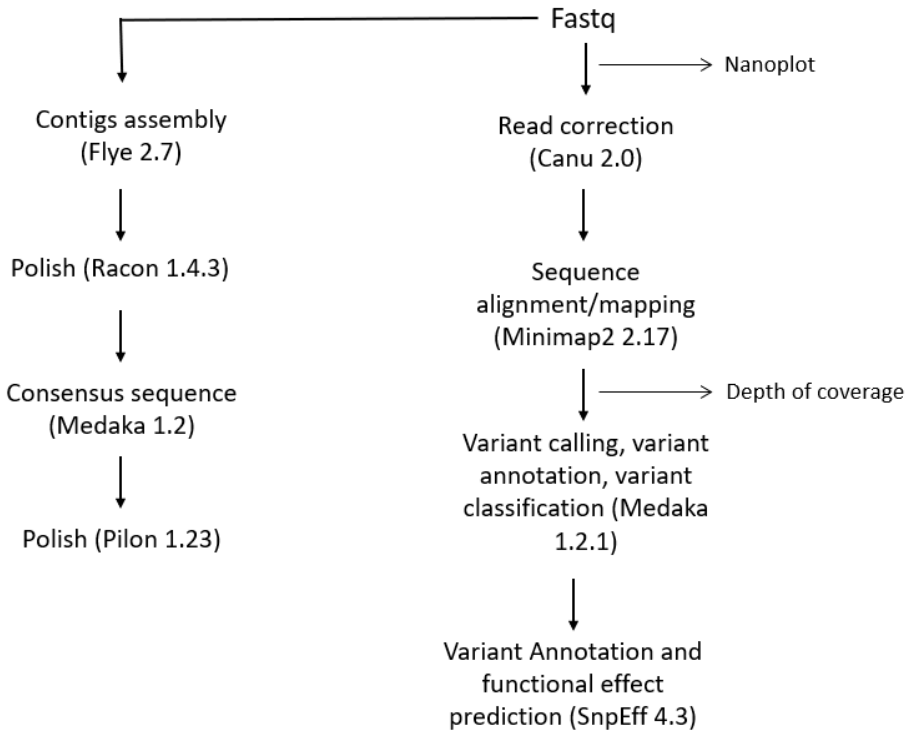


Figure 4. Flow diagram with bioinformatics analysis pipelines for SNP analysis (right) and for genome assemblies for MLST analysis (left).

5. Multi Locus Sequence Typing

Using *C. albicans*[58] and *C. glabrata*[80] MLST schemes hosted in PubMLST, the assembled genomes of each sample were compared to the reference database in order to identify allelic matches and define a strain type. Results were ordered by best match. If no exact match was found for all genes defined in the MLST gene set, the locus/scheme was selected using the closest match for the gene(s) lacking.

6. Genomic variant analysis

Genomic variant calling, annotation and classification were done using the appropriate Medaka (v.1.2.1) pipeline for: (i) identification of variants, (ii) annotation of the vcf file with read depth and supporting reads for the variant calling, and (iii) classification of variants. SnpEff (v4.3)[81] was then used to annotate variants based on their genomic locations (intronic, untranslated region,

upstream, downstream, splice site, or intergenic regions) and predict functional effect of variants (synonymous or non-synonymous amino acid replacement, start codon gains or losses, or frame shifts)[81]. The pipeline for variant calling and annotation is described in Figure 4 (right).

Quality filtering on the annotated variants was then performed; all low-quality annotations were excluded; only quality scores for alternative basecalling ≥ 10 were kept, assuming a Phred quality score of 10%. Only variants with at least one read spanning position 25, and with depth of reads well aligned to the allele ≥ 10, were kept.

7. Gene Ontology analysis

A Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to explore (i) genes with common SNPs among isolates with distinct origin (vaginal exudate, oral cavity and haemoculture), and (ii) genes with unique SNPs from isolates collected from the same type of sample (vaginal exudate, oral cavity and haemoculture). To identify the biological processes, molecular functions, and cellular components in which these groups of genes were involved, we performed gene ontology (GO) analysis. For this, the GO Slim mapper (http://www.candidagenome.org/cgi-bin/GO/goTermMapper) and the Go Term finder (http://www.candidagenome.org/cgi-bin/GO/goTermFinder) tools available at the Candida Genome Database[18] were used to annotate the genes according to each of the GO sets in broader groups and to find significant shared GO terms in each GO set to describe the genes SNPs were found (p-value ≤ 0.1), respectively.

8. Antifungal susceptibility testing

8.1. Antifungal drugs, control strains and culture media

Susceptibility of the isolates in study and minimum inhibitory concentrations (MIC) of fluconazole (FL), amphotericin B (AP), anidulafungin (AND), micafungin (MYC) and caspofungin (CS), were determined using Etest (bioMérieux, France) and according to EUCAST guidelines[82], and the recommendations of the Etest package insert (9305056D - en - 2013/02). Control strains used to monitor tests performance were *C. albicans* SC5314, and *C. albicans* T0, which parental strain is *C. albicans* SN148 wild-type, which were previously shown to be susceptible to fluconazole[83], kindly provided by Dr. Ana Rita Bezerra from Instituto de Biomedicina (iBiMED), Universidade de Aveiro.

YPD (Formedium) solid medium with 2% agar was used to subculture fungal isolates, except for strain T0, which was grown in synthetic defined (SD) medium without uracil containing 0.67% yeast nitrogen base, 2% glucose, 0.2% of dropout mix and 2% agar, and in RPMI 1640 solid medium

with 2% glucose. RPMI 1640 solid medium (Sigma-Aldrich) with 2% glucose was used for the antifungal susceptibility tests.

8.2. Etest MIC and susceptibility determination

The *Candida* isolates in study and control strains were inoculated on YPD, and incubated at 35°C for 24 hours. Yeast colonies (approximately > 1 mm diameter) on YPD were suspended in 0.9% sterile saline solution to adjust to a turbidity of a 0.5 McFarland standard. A micropipette and a loop were used to spread 125 µl of fungal cells suspension evenly on the RPMI plates. The Etest strips were placed on the plates after they dried for approximately 15 minutes at room temperature. The strip end with a lower concentration of the antifungal was positioned first. The concentration of antifungal on the strip ranged from 0.002 to 32 µg/ml (AND, AP, MYC, CS) and 0.016 to 256 µg/ml (FL). Control strain T0 was only tested for susceptibility to FL.

The MICs of the antifungal drugs were read after incubation for 24 and 48 hours at 35°C, where the growth inhibition ellipse edge intersected the strip. Regarding MIC determination for AP, the MIC value was read at complete inhibition of growth; for AND, MYC and CS, the reading of MIC was done considering trailing endpoints at the first visual point of significant inhibition; for FL, the MIC was read at the first point of significant inhibition or marked decrease in growth density (usually at 80% growth inhibition). Interpretation of results for determining the isolates susceptibility profiles was done according to the EUCAST guidelines[82].

8.3. Broth microdilution assay for determining MIC of fluconazole and isolates susceptibility

The MIC of FL was additionally determined by broth microdilution, according to the EUCAST guidelines. Essentially, the isolates were inoculated on YPD and incubated at 35°C for 24 hours. The inoculum was prepared by suspending yeast colonies (approximately > 1 mm diameter) in 0.9% sterile saline solution corresponding to a 0.5 McFarland standard, and adjusted to a final concentration of 1-5x10^6 cells/ml. The fluconazole powder (Sigma-Aldrich) was dissolved in dimethylsulfoxide (DMSO) to a concentration of 5000 µg/ml. The solution was then diluted in 2x RPMI 1640 medium (Sigma-Aldrich) with 2% glucose, buffered to pH 7.0 with 0.165 M morpholinepropanesulfonic acid (MOPS) buffer (Sigma), sterilized by filtration. Both inoculum and fluconazole solutions were distributed in sterile flat bottom 96-well microplates (Corning® Life Sciences) reaching final drug concentrations ranging from 0.125 to 256 µg/ml and 4.0x10^5 CFU/ml of cell culture (Figure 5). After incubation without agitation for 24 hours at 35°C, the microplates were

read using a microdilution plate reader at a wavelength of 595 nm, shortly after resuspension of cells with a multi-channel pipette. Each microplate was read three times.

For quality control purposes, *C. parapsilosis* ATCC 22019 was used in the assays.

The MIC was defined as the lowest drug concentration that resulted in a reduction in growth by 50% or more of that of the fluconazole-free control[69].

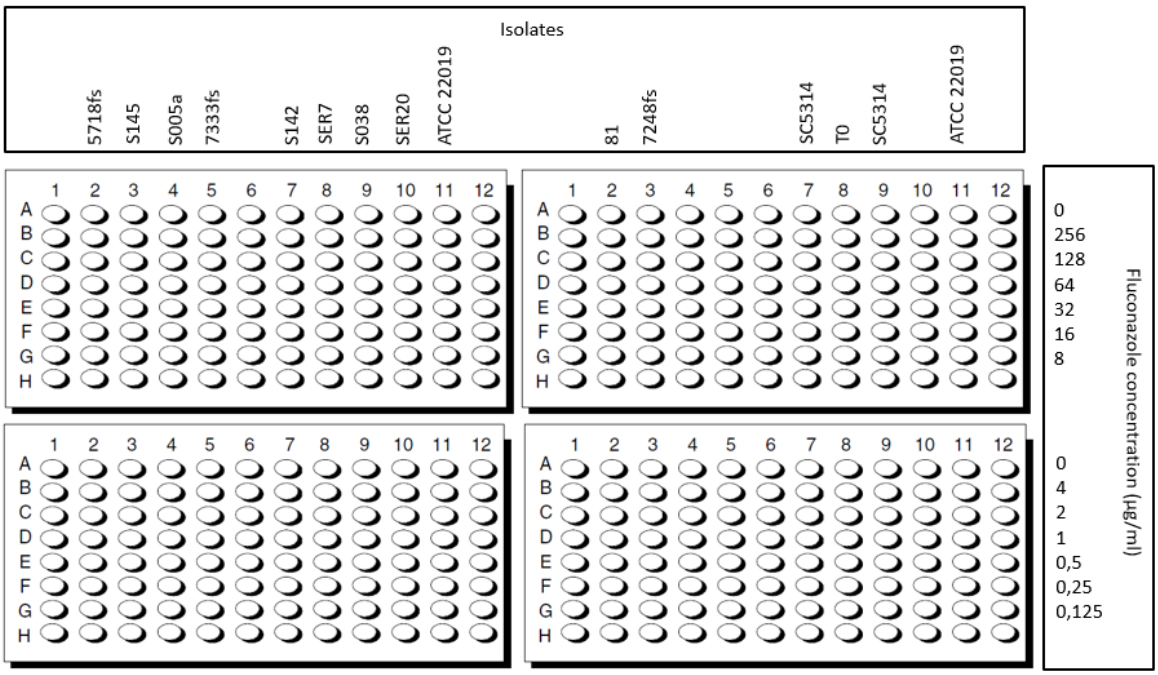


Figure 5. Fluconazole concentrations and isolates distribution on microplates.

III. RESULTS

1. Whole-Genome Sequencing

1.1. DNA quantification and genomic library preparation

During the preparation of the genomic library, DNA quantification using the Qubit assay was performed several times, according to the protocol mentioned before. DNA concentrations obtained for each sample ranged from 804 ng/µl to 2890 ng/µl in the first run, and 230 ng/µl to 790 ng/µl in the second run. Final library concentrations which were loaded for sequencing were 10 ng/µl and 24.4 ng/µl, for the first and second run, respectively.

1.2. Sequencing run metrics

In total, 1358500 reads were obtained from the two sequencing runs, ranging from 53649 for barcode 09 (sample S038) and 396547 for barcode 07 (sample 81) and a mean read length of 6.5 kb. Mean read quality ranged from 12.2 for barcode 02 (sample SER7) and 12.9 for barcode 08 (sample 7333fs). Table 4 depicts statistics per sample.

Table 4. Summary statistics of run metrics, per sample.

Isolate	SER20	SER7	S142	S145	S005a
Number of reads	61658	79264	59448	70858	102874
Mean read length (bp)	5942.3	8369.9	6880.4	6210.6	6243.8
Maximum read size (bp)	156709	123760	101132	108600	116290
Read length N50 (bp)	9699	14578	12080	10295	10707
Mean read quality	12.5	12.2	12.3	12.8	12.5
Isolate	5718fs	81	7333fs	S038	7248fs
Number of reads	163951	396547	265192	53649	105059
Mean read length (bp)	5936.1	5645.6	7091.1	6261.5	6444.1
Maximum read size (bp)	103569	126774	119838	78079	140296
Read length N50 (bp)	9669	9004	11838	10382	10697
Mean read quality	12.6	12.6	12.9	12.3	12.7

2. Bioinformatic analysis

After read correction with Canu to improve the accuracy of bases in reads, they were mapped against the *C. albicans* reference genome SC5314 haplotype A (Table 5). The number of total reads ranged from 30791 for sample 7333fs and 65949 for sample S145; the percentage of reads mapped against *C. albicans* reference genome ranged between 38.9% for sample SER7 and 99.7% for sample S005a.

Table 6 shows the depth of coverage at chromosome level and whole genome level in each *Candida* isolate. Sequencing depth of coverage at the genome level ranged from 1.54 for sample SER7 and 37.32 for sample 81.

Table 5. Mapping characteristics per sample.

Isolate	SER20	SER7	S142	S145	S005a
Genome length (Mb)	14.3	14.3	14.3	14.3	14.3
Total reads	60548	4180.5	55394	65949	64770
Total reads mapped	60307	1627.1	55135	65659	64584
Percentage of reads mapped	99.6	38.9	99.5	99.6	99.7
Isolate	5718fs	81	7333fs	S038	7248fs
Genome length (Mb)	14.3	14.3	14.3	14.3	14.3
Total reads	39906	44700	30791	52616	56494
Total reads mapped	39292	38940	25105	52253	56196
Percentage of reads mapped	98.5	87.1	81.5	99.3	99.5

Table 6. Sequencing depth of coverage for each chromosome in each isolate.

Chromosome	Length (bp)	SER20	SER7	S142	S145	S005a
ca22chr1a	3188341	22.90	1.1	26.1	28.2	34
ca22chr2a	2231883	23.30	0.9	26.5	27.9	34.9
ca22chr3a	1799298	23.10	0.8	26.4	27.8	35.6
ca22chr4a	1603259	22.80	0.7	26.4	27.9	35.5
ca22chr5a	1190869	21.30	0.7	25.4	27.2	32.1
ca22chr6a	1033292	21.80	0.6	25.5	26.4	36.3
ca22chr7a	949580	24.30	0.7	25.5	28.1	36.8
ca22chrm	40420	163.10	6.6	211.4	131.5	260.7
ca22chrra	2286237	27.50	5.1	29.2	32.7	42.5
whole genome	143233179	23.99	1.54	27.11	28.86	36.70
Chromosome	Length (bp)	5718fs	81	7333fs	S038	7248fs
ca22chr1a	3188341	34.2	33.2	35.1	21.1	34.4

ca22chr2a	2231883	34.3	34.8	36.2	21.3	35.3
ca22chr3a	1799298	37	36.8	35.1	21.1	35.1
ca22chr4a	1603259	35.6	35.7	36.1	21.1	35.1
ca22chr5a	1190869	34.4	34.5	34.9	20.4	34.2
ca22chr6a	1033292	33.8	35	34.9	21.2	34.6
ca22chr7a	949580	37.9	36.9	34.6	22.8	37.9
ca22chrm	40420	270.5	352.7	174	157.9	273.2
ca22chrra	2286237	41.3	44.2	43.1	24.4	42.3
whole genome	143233179	36.76	37.32	36.99	22.11	36.87

2.2. Multi Locus Sequence Typing

The results of the multi locus sequence typing for *C. albicans* revealed that for some locus, the alleles of the isolates were not present in the PubMLST database and a definitive allocation to a sequence type was not possible for any sample. For samples S005a, 5718fs, S038, over 20 profiles closely matching to a previously described ST were found and are not all depicted in the table. The relevant data was submitted to the *Candida albicans* PUBMLST database for the definitive assignment of STs to these isolates. Table 7 shows the identified and closest matches to the alleles previously known and the STs closest matches for each sample. Because a definitive allocation to a sequence type was not possible for any sample, we were not able to determine to which clade each isolate belonged to.

Table 7. MLST results for *C. albicans*. The alleles referenced with * correspond to the closest match found for the respective locus.

Sample	Locus	AAT1a	ACC1	ADP1	MPIb	SYA1	VPS13	ZWF1b	Nearest match ST
SER20	Allele	3*	58	10*	5	34*	20	20	790; 1493; 3402
SER7		-	-	-	-	-	-	239*	-
S142		8*	2	6	9*	69*	20*	161*	1150; 2745
S145		8*	3	30*	4*	34*	13*	276	1603; 2579
S005a		33*	58	10*	5	34	58	159*	215; 790; 797
5718fs		8*	2	2	9*	2*	207*	20*	54; 76; 261
81		33*	52*	10	4	34	4	196*	1483
7333fs		3*	3	2	9*	2*	24	196	310

S038		107*	58	6	4	34*	54	205*	167; 287; 441
7248fs		13*	8	10	19*	53*	13	205*	1485

2.3. Identification of isolate SER7

As depicted in Table 5, 38.9% of SER7 sequencing reads were mapped against *C. albicans* reference genome (version A22-s07-m01-r100) and the sequencing depth of coverage was only 1.54x, not following the trend seen for the remaining samples. Additionally, querying the ST of this isolate in the *C. albicans* PubMLST database produced only one closest match for gene *ZWF1b.* To ascertain this isolate species identification, a blastn search against the NCBI nucleotide database of this isolate genome (2020.11.17) was performed and results showed closest match to *C. glabrata* (Figure 6).

SER7 reads were then mapped against the *C. glabrata* reference genome CBS138 (version s03-m01-r04) (Table 8). The number of total reads was 48504, with a percentage of reads mapped against *C. glabrata* reference genome of 99.7% and an overall sequencing depth of coverage of 42.7x (Table 9).

BLASTN 2.10.1+

Reference:
Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000),
"A greedy algorithm for aligning DNA sequences", J Comput Biol 2000;
7(1-2):203-14.

Database: Nucleotide collection (nt)
60,828,507 sequences; 327,482,466,005 total letters

Query= contig_1

Length=1302913

Sequences producing significant alignments:	Score (Bits)	E Value
CP048240.1 [Candida] glabrata strain BG2 chromosome K	4.729e+05	0.0
CP048128.1 [Candida] glabrata strain ATCC 2001 chromosome K	4.390e+05	0.0
CR380957.2 Candida glabrata strain CBS138 chromosome K complete s...	4.390e+05	0.0
XM_448582.1 [Candida] glabrata uncharacterized protein (CAGL0K082...	22232	0.0
XM_448497.1 [Candida] glabrata uncharacterized protein (CAGL0K063...	17872	0.0
XM_448247.1 [Candida] glabrata uncharacterized protein (CAGL0K005...	13592	0.0
XM_448671.1 [Candida] glabrata uncharacterized protein (CAGL0K103...	11273	0.0
HM366442.1 Candida glabrata strain ATCC 90030 beta-1,3-glucan syn...	10510	0.0
XM_448401.1 [Candida] glabrata uncharacterized protein (CAGL0K040...	10504	0.0
HM366441.1 Candida glabrata strain 4872 truncated beta-1,3-glucan...	10432	0.0

Figure 6. Results of blastn search against the NCBI nucleotide database of SER7 genome.

Table 8. Metrics for SER7 sequencing reads mapping against *C. glabrata* CBS138 reference genome (version s03-m01-r04).

Isolate	SER7
Genome length (Mb)	12.3
Total reads	48504
Total reads mapped	48381
Percentage of reads mapped	99.7

Table 9. Sequencing depth of coverage for each chromosome in SER7.

Chromosome	Length (bp)	SER7
chra_c_glabrata_cbs138	491328	49.8
chrb_c_glabrata_cbs138	502101	43.0
chrc_c_glabrata_cbs138	558804	41.9
chrd_c_glabrata_cbs138	651701	42.1
chre_c_glabrata_cbs138	687738	40.6
chrf_c_glabrata_cbs138	927101	37.8
chrg_c_glabrata_cbs138	992211	39.1
chrh_c_glabrata_cbs138	1050361	39.3
chri_c_glabrata_cbs138	1100349	42.3
chrj_c_glabrata_cbs138	1195129	39.5
chrk_c_glabrata_cbs138	1302831	39.6
chrl_c_glabrata_cbs138	1455689	52.1
chrm_c_glabrata_cbs138	1402898	38.7
mito_c_glabrata_cbs138	20063	568.9
whole genome	12338304	42.7

The results of the multi locus sequence typing analysis for SER7 revealed this is a *C. glabrata* ST 10 isolate (Table 10), which belongs to group 3[61].

Table 10. MLST results for *C. glabrata.*

Sample	Locus	FKS	LEU2	NMT1	TRP1	UGP1	URA3	ST
SER7	Allele	8	4	3	5	1	2	10

3. Genomic variants

A total of 1292088 SNPs were called with the methods used herein. After applying the quality filters described in the methods, a total of 791021 SNPs were found, ranging from 50320 in sample S142 to 104090 in sample 81 (Table 11).

Table 11. Number of variants called per sample.

Sample	Number of SNPs called	Number of SNPs called after filtering
SER20	144341	85604
S142	95102	50320
S145	154672	96333
S005a	150313	96943
5718fs	95167	54074
81	163155	104090
7333fs	114065	65407
S038	143900	82051
7248fs	149619	96754
SER7	81754	59445

From the total of SNPs filtered and included in this analysis, 380908 were located in noncoding regions, and 410113 were located in coding regions. Among these, 135641 were missense and 255303 were synonymous variants. The number of SNPs found in noncoding regions ranged from 13889 (approximately 23%) in sample SER7 to 52357 (approximately 50%) in sample 81. The number of missense SNPs ranged from 8565 (approximately 14%) in sample SER7 to 17854 (approximately 17%) in sample 81. The number of synonymous SNPs ranged from 15940 (approximately 32%) in sample S142 to 33483 (approximately 32%) in sample 81. Figure 7 shows the percentage of different SNPs identified per sample. There was clear difference in the pattern of SNPs in noncoding regions in *C. albicans* isolates relative to the *C. glabrata* isolate, SER7. The latter had approximately 23% of SNPs in noncoding regions, while *C. albicans*, on average, had approximately 50%. All *Candida* isolates seemed to follow the same pattern regarding the synonymous SNPs; for the missense SNPs, all *C. albicans* isolates followed the same pattern, having on average around 17% of missense SNPs, while the *C. glabrata* isolate, SER7, had a little less, approximately 14% of missense SNPs.

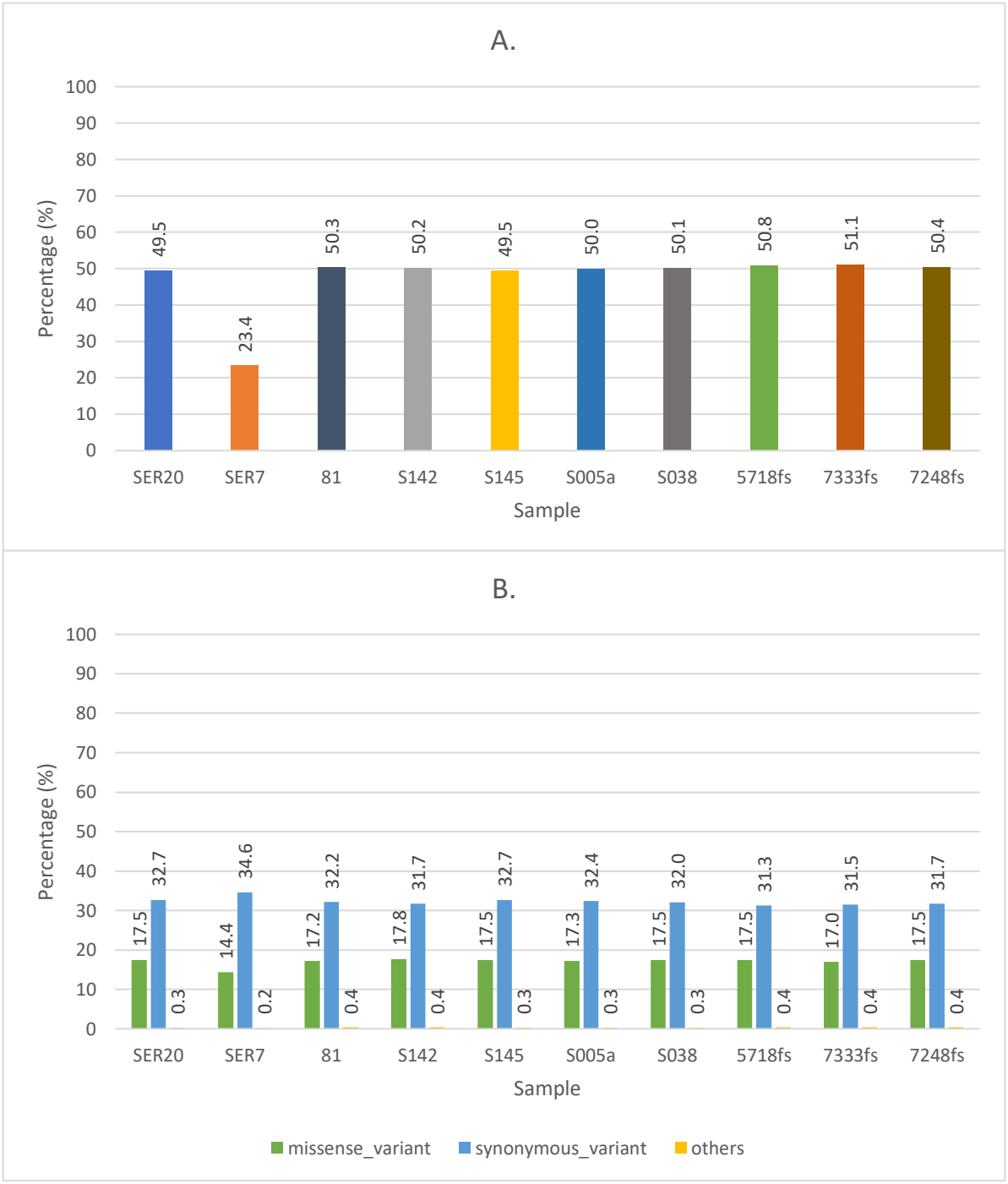


Figure 7. Distribution of SNPs found in A. noncoding and B. coding regions, per sample.

According to their source of collection, *C. albicans* samples showed also different missense SNP amounts. Samples from the oral cavity (S142, S145, S005a, and S038) registered the higher frequency of missense SNPs, with 56775 (8%) SNPs found, followed by samples from vaginal exudates (5718fs, 7333fs, and 7248fs) and haemoculture (SER20 and 81), with a total of 37477 (5%) and 32824 (5%) missense SNPs, respectively. Figure 8 shows the distribution of missense SNPs according to sample origin.

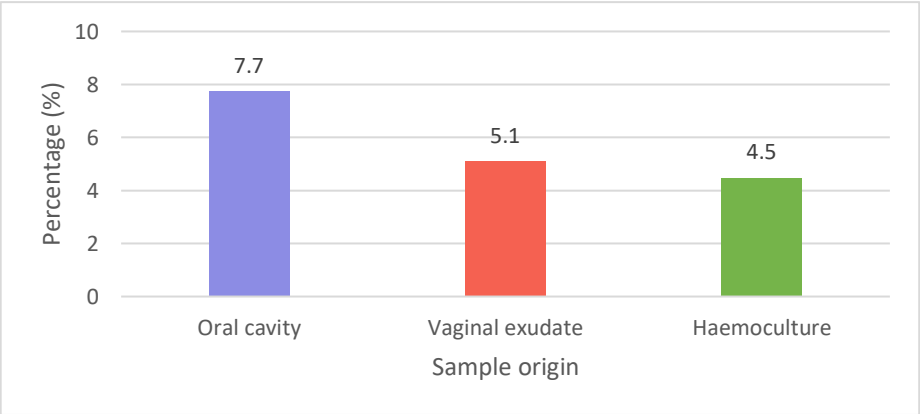


Figure 8. Distribution of missense SNPs found per type of sample, regarding their origin, in relation to the total number of SNPs found.

In total, 664 genes with missense SNPs were found to be common to isolates from different sources and 41, 239, and 1751 were unique to samples from the oral cavity, vaginal exudates and haemoculture, respectively (Figure 9).

SNPs found in the isolates were not the ones described in the literature to be associated to antifungal resistance. However other SNPs were found in those genes, but with no relation to resistance phenotype predicted.

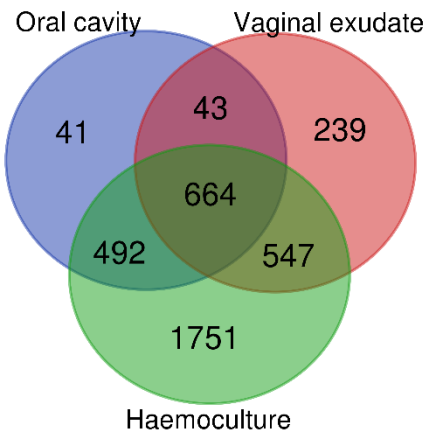


Figure 9. Venn diagram displaying the number of common and unique sets of genes with missense SNPs between samples from the three different sources.

3.1. GO analysis of *C. albicans* isolates

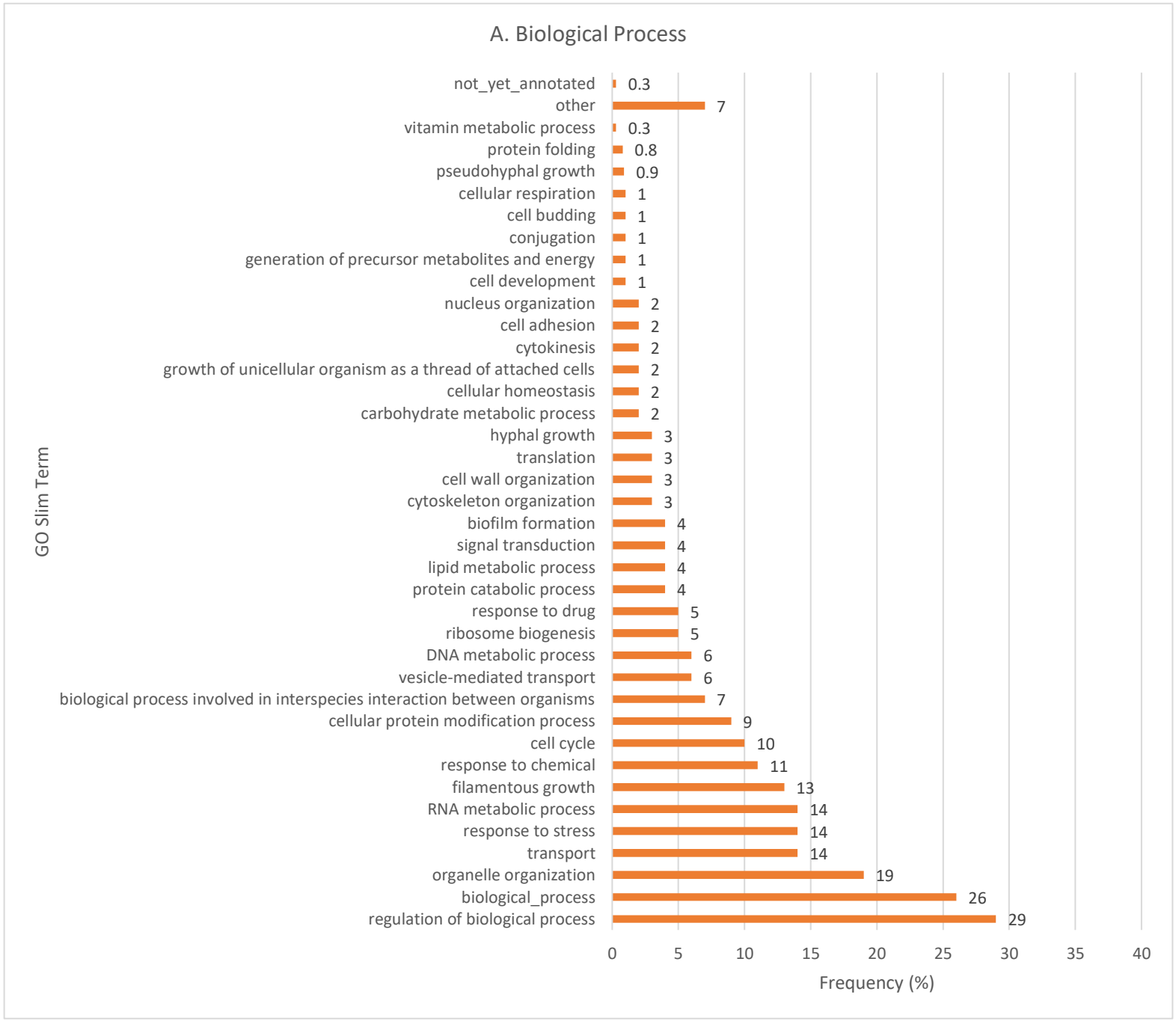
To look for a possible role of the genes with missense SNPs at the level of molecular function, biological process, or cellular component, using GO Slim mapper and GO Term finder were used. We also explored similarities and differences between mutated genes in samples from common and distinct sources using GO annotations.

3.1.1. Genes with SNPs common to *C. albicans* isolates from different types of sources

Among genes with SNPs common to all samples (n=664), (i) 38% mapped against genes coding for cellular components, from which cytoplasm (32%), nucleous (25%), and membrane (17%) were the most common; (ii) 29% were genes encoding products with a role in the regulation of biological processes, principally, organelle organization (19%), transport (14%), and response to stress (14%); and 36% were genes coding for products involved in molecular-level activities, mainly hydrolase activity (16%), transferase activity (12%), and DNA binding (9%). Figure 10 depicts the frequencies of *C. albicans* genes with missense SNPs that are common to the three types of samples (haemoculture, vaginal exudates, and oral cavity) annotated by GO Slim terms.

The GO Term for biological process most associated to genes with SNPs common to all samples (n=664), was adhesion of symbiont to host (p-value 0.026; specific genes were *ace2, als1-4, hwp2, phr1, rfx2, sap2, sap3, sap6, ura3,* and *ywp1*). The terms negative regulation of macromolecule metabolic process (p-value 0.061), and negative regulation of nucleobase-containing compound metabolic process (p-value 0.090) were not significantly associated to genes with SNPs common to all samples. Among the molecular function set, transcription regulator activity (p-value 0.057) and DNA-binding transcription factor activity (p-value 0.061) were the GO terms found, although not with statistic significance; these two GO Terms have a parent-child relationship, as DNA-binding transcription factor activity is a transcription regulator activity[84]. GO terms of the cellular components background associated to this set of genes were not found.

Figure 11 shows GO Terms associated to genes with SNPs common to all samples, and Appendix 1 contains the graphic displays with the relationships among the GO terms found for this gene set and the GO terms to which genes were directly annotated.


A. Biological Process
GO Slim Term
not_yet_annotated 0.3
other 7
vitamin metabolic process 0.3
protein folding 0.8
pseudohyphal growth 0.9
cellular respiration 1
cell budding 1
conjugation 1
generation of precursor metabolites and energy 1
cell development 1
nucleus organization 2
cell adhesion 2
cytokinesis 2
growth of unicellular organism as a thread of attached cells 2
cellular homeostasis 2
carbohydrate metabolic process 2
hyphal growth 3
translation 3
cell wall organization 3
cytoskeleton organization 3
biofilm formation 4
signal transduction 4
lipid metabolic process 4
protein catabolic process 4
response to drug 5
ribosome biogenesis 5
DNA metabolic process 6
vesicle-mediated transport 6
biological process involved in interspecies interaction between organisms 7
cellular protein modification process 9
cell cycle 10
response to chemical 11
filamentous growth 13
RNA metabolic process 14
response to stress 14
transport 14
organelle organization 19
biological_process 26
regulation of biological process 29
0 5 10 15 20 25 30 35 40
Frequency (%)

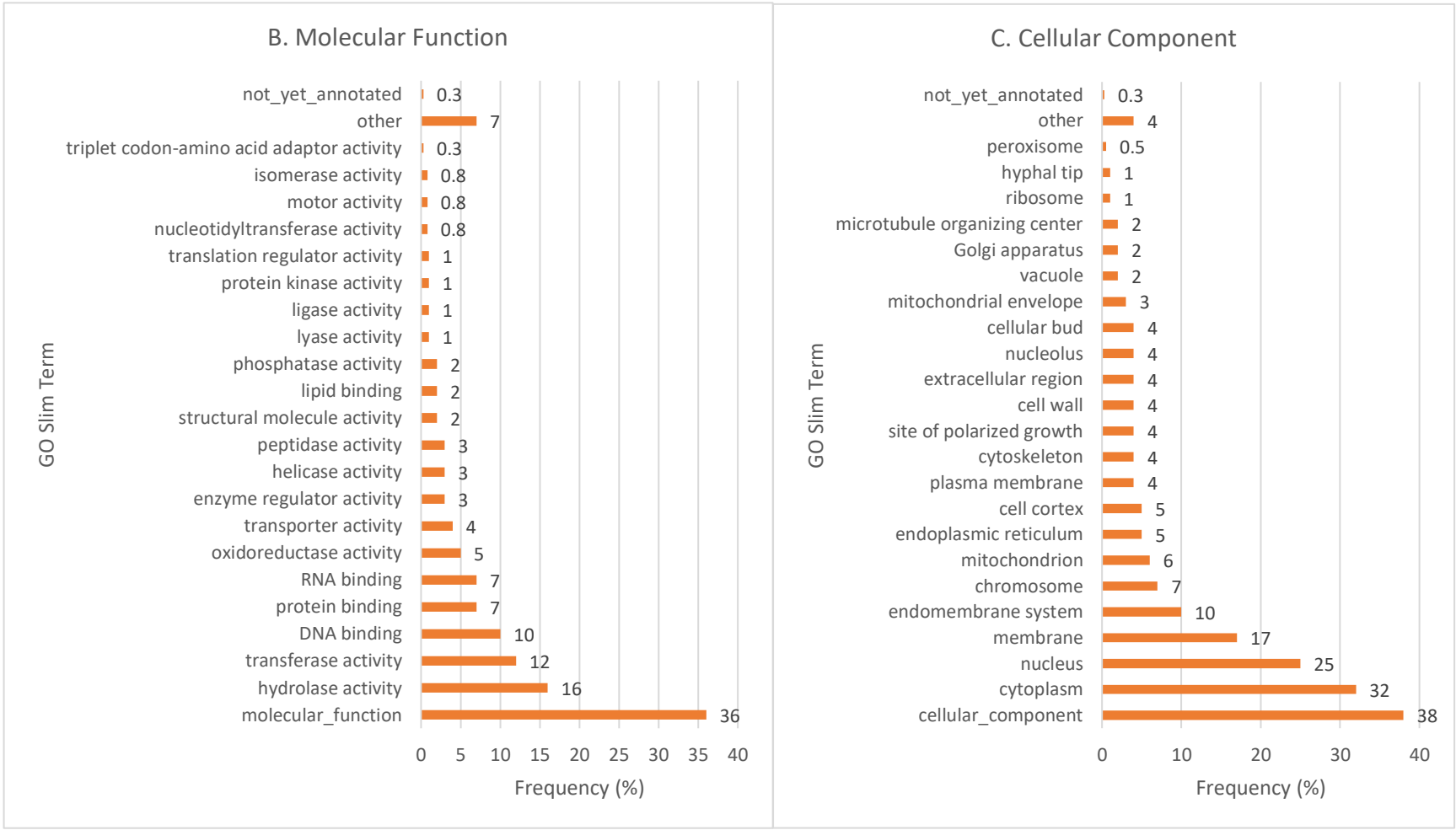


Figure 10. Frequencies of GO Slim terms mapped to genes with missense SNPs that are common to the three types of sample. A. GO Slim terms referent to Biological Process; B. GO Slim terms referent to Molecular Function; C. GO Slim terms referent to Cellular Component.

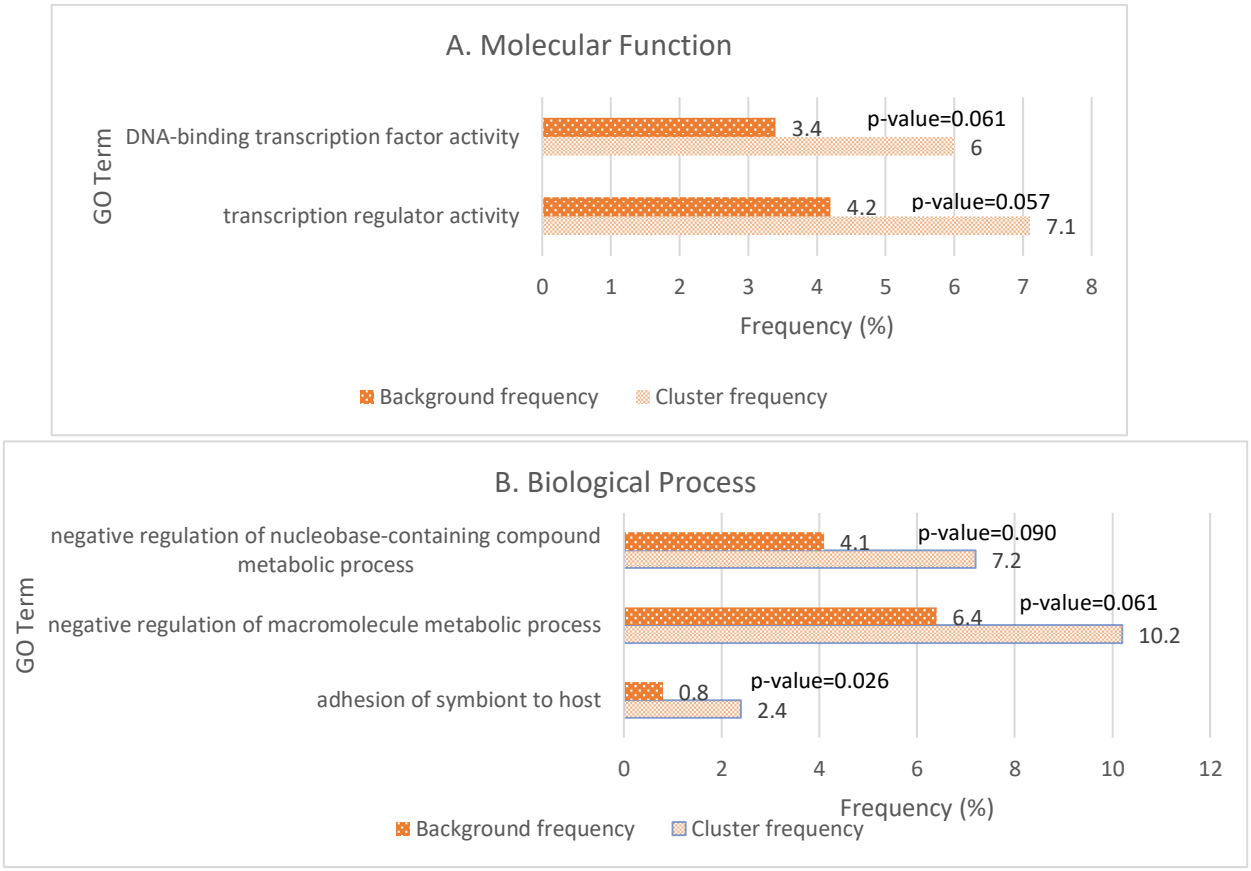


Figure 11. GO Term cluster frequencies of genes with missense SNPs common to isolates from distinct sources. A. GO Terms referent to Molecular Function; B. GO Terms referent to Biological Process.

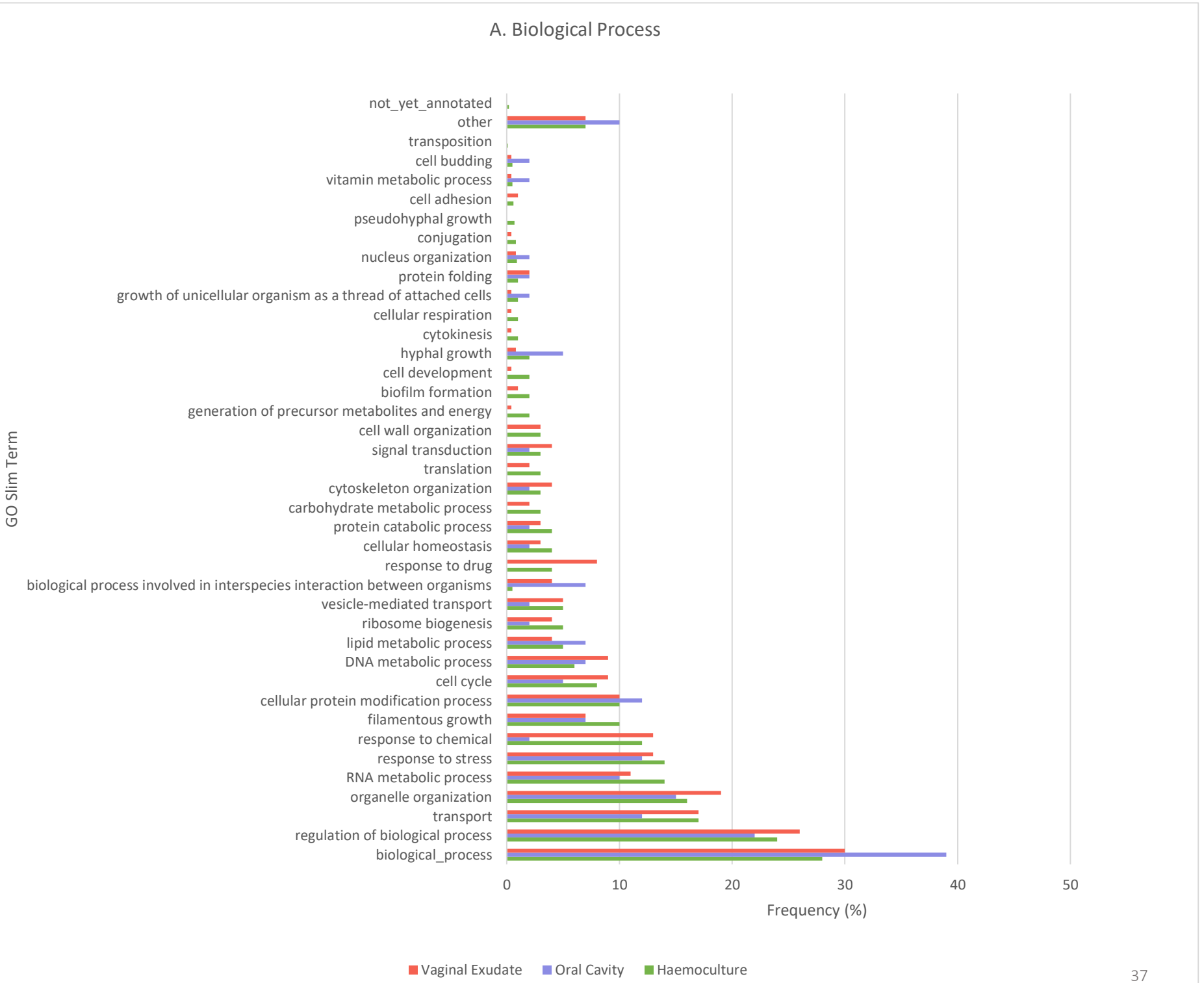
3.1.2. Genes with SNPs unique to *C. albicans* isolates from different types of sources

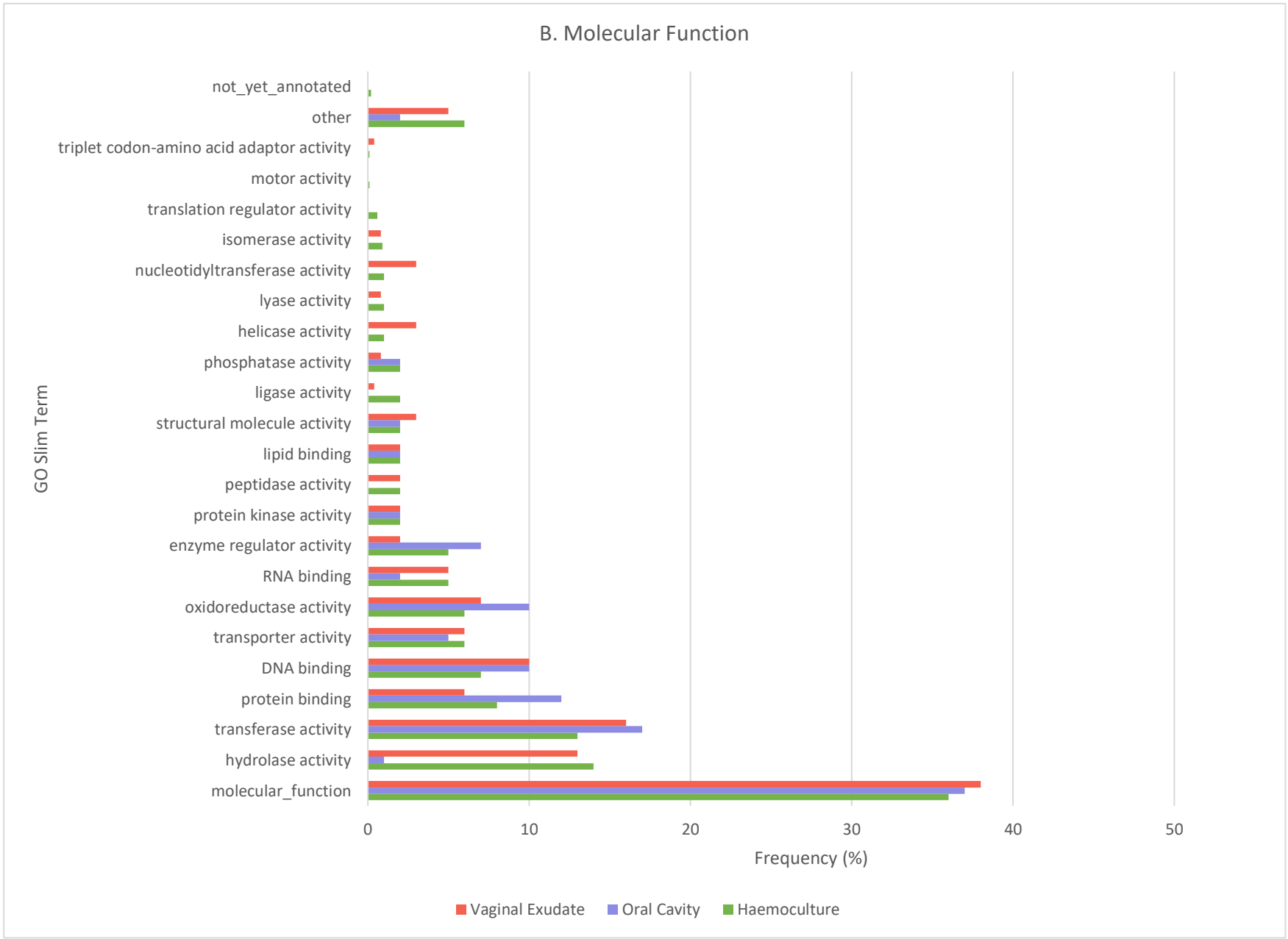
Among genes with unique SNPs found in isolates from haemoculture samples, (n=1751), vaginal exudate samples (n=239), and oral cavity samples (n=41), (i) 38%, 42%, and 51% mapped against genes coding for cellular components, respectively. The most common were cytoplasm, nucleous and membrane for all isolates from different types of sources, with 33%, 23%, and 20% for haemoculture samples, 31%, 22%, and 18% for vaginal exudate samples, and 32%, 24%, and 17% for oral cavity samples; (ii) 28%, 30%, and 39% were genes encoding products with a role in biological processes, found in haemoculture, vaginal exudate, and oral cavity samples, respectively; mainly regulation of biological process (24%, 26%, and 22%), organelle organization (16%, 19%, and 15%), and transport (17%, 17%, and 12%), for samples from haemoculture, vaginal exudate, and oral cavity, respectively; and 36%, 38%, and 37% were genes coding for products involved in molecular-level activities, for haemoculture, vaginal exudate, and oral cavity samples; the most common for haemoculture and vaginal exudates samples were hydrolase activity (14% and 13%, respectively), and transferase activity (13% and 16%, respectively); for samples originated from the oral cavity, the

most common were transferase activity (17%) and protein binding (12%). Figure 12 depicts the frequencies of *C. albicans* genes with missense SNPs that are unique to each type of samples (haemoculture, vaginal exudates, and oral cavity) annotated to GO Slim terms. The great majority of genes with unique missense SNPs found in haemoculture samples, vaginal exudate samples, and oral cavity samples, mapped against genes coding for cellular components, molecular functions, and biological processes, did not show marked variance. The exception was in samples obtained from the oral cavity, concerning the absence of genes encoding cellular components, namely plasma membrane, site of polarized growth, vacuole, extracellular region, and cell cortex, which did not happen for haemoculture and vaginal exudate samples; in genes coding for products involved in molecular-level functions there were differences in their frequency among the different types of isolates, namely in hydrolase activity (1% for oral cavity samples, 14% for haemoculture samples, and 13% for vaginal exudate samples), in protein binding (12% for oral cavity samples, 8% for haemoculture samples, and 6% for vaginal exudate samples); and in genes encoding for biological processes, such as response to chemical (2% for oral cavity samples, 12% for haemoculture samples, and 13% for vaginal exudate samples), and in hyphal growth (5% for oral cavity isolates, 2% for haemoculture isolates, and 0.8% for vaginal exudate isolates).

GO Terms within the biological process set significantly associated to genes with SNPs unique to samples obtained from the oral cavity (n=41), was oligosaccharide-lipid intermediate biosynthetic process (p-value 0.04; genes *alg5, alg8,* and *ecm39*). For samples originated from vaginal exudates, among the molecular function set, lysophospholipid acyltransferase activity (p-value 0.04; genes *c1_04110w_a, c3_00400c_a,* and *lpt1*) was the GO Term significantly associated to genes with SNPs unique to this group of samples; transferase activity, transferring acyl groups other than amino-acyl groups (p-value 0.06) and lysophospholipid acyltransferase activity (0.09) were not significantly associated to genes with SNPs unique to vaginal exudate samples. These GO Terms have a parent-child relationship, as lysophosphatidic acid acyltransferase activity is a lysophospholipid acyltransferase activity, which in turn is a transferase activity, transferring acyl groups other than amino-acyl groups[84]. GO terms of the cellular components background associated to these set of genes (unique to samples from different sources) were not found.

Figure 13 shows GO Terms associated to genes with SNPs unique to samples from different backgrounds, and Appendix 2 and 3 contain the graphic displays with the relationships among the GO terms found for these gene sets and the GO terms to which genes were directly annotated.


A. Biological Process
GO Slim Term
not_yet_annotated
other
transposition
cell budding
vitamin metabolic process
cell adhesion
pseudohyphal growth
conjugation
nucleus organization
protein folding
growth of unicellular organism as a thread of attached cells
cellular respiration
cytokinesis
hyphal growth
cell development
biofilm formation
generation of precursor metabolites and energy
cell wall organization
signal transduction
translation
cytoskeleton organization
carbohydrate metabolic process
protein catabolic process
cellular homeostasis
response to drug
biological process involved in interspecies interaction between organisms
vesicle-mediated transport
ribosome biogenesis
lipid metabolic process
DNA metabolic process
cell cycle
cellular protein modification process
filamentous growth
response to chemical
response to stress
RNA metabolic process
organelle organization
transport
regulation of biological process
biological_process
0
10
20
30
40
50
Frequency (%)
Vaginal Exudate
Oral Cavity
Haemoculture


B. Molecular Function
not_yet_annotated
other
triplet codon-amino acid adaptor activity
motor activity
translation regulator activity
isomerase activity
nucleotidyltransferase activity
lyase activity
helicase activity
phosphatase activity
ligase activity
structural molecule activity
lipid binding
peptidase activity
protein kinase activity
enzyme regulator activity
RNA binding
oxidoreductase activity
transporter activity
DNA binding
protein binding
transferase activity
hydrolase activity
molecular_function
GO Slim Term
0
10
20
30
40
50
Frequency (%)
Vaginal Exudate
Oral Cavity
Haemoculture

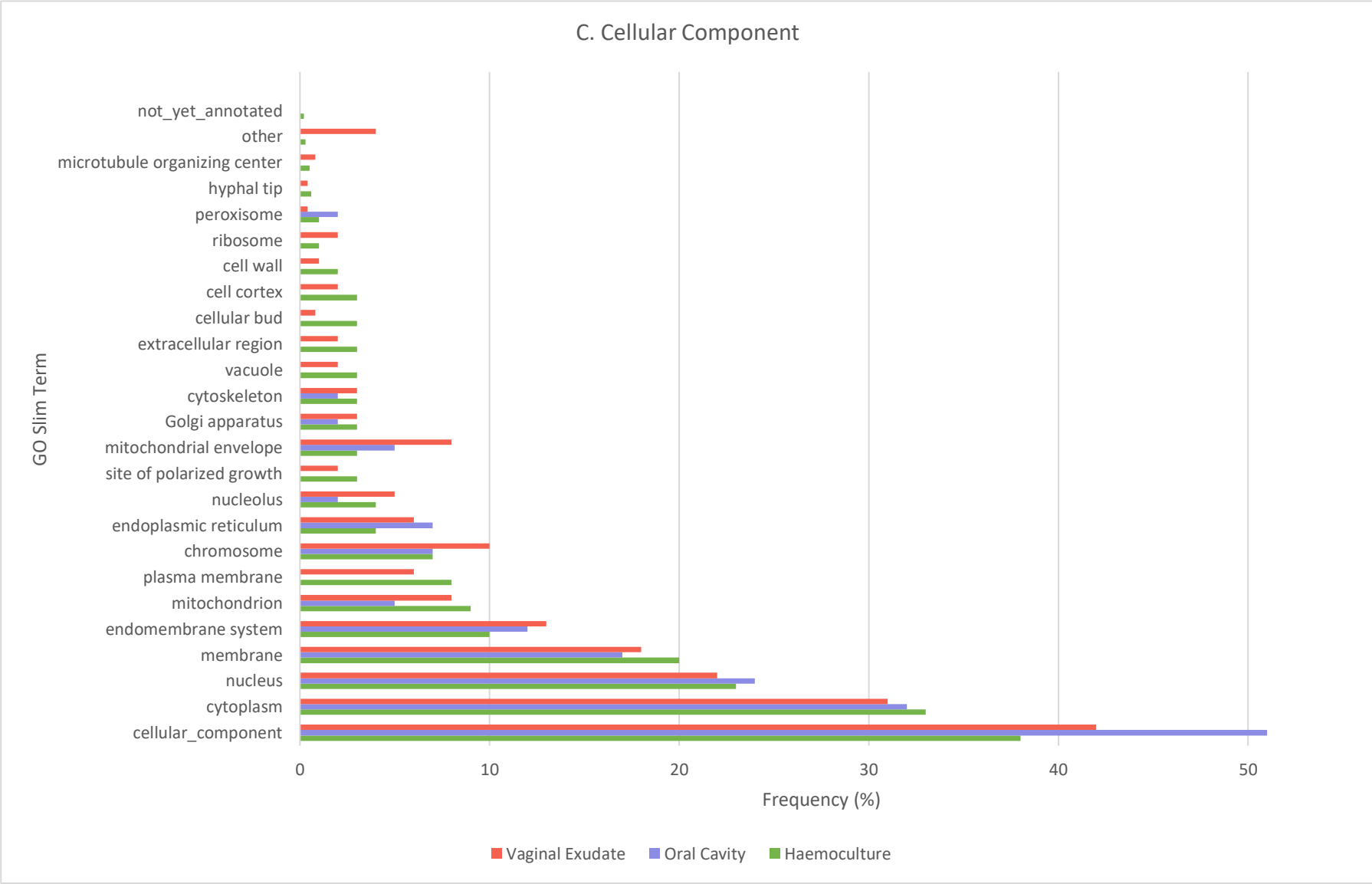


Figure 12. GO Slim terms frequencies in *C. albicans* genes with unique missense SNPs, according to sample origin. A. GO Slim terms referent to Biological Process; B. GO Slim terms referent to Molecular Function; C. GO Slim terms referent to Cellular Component.

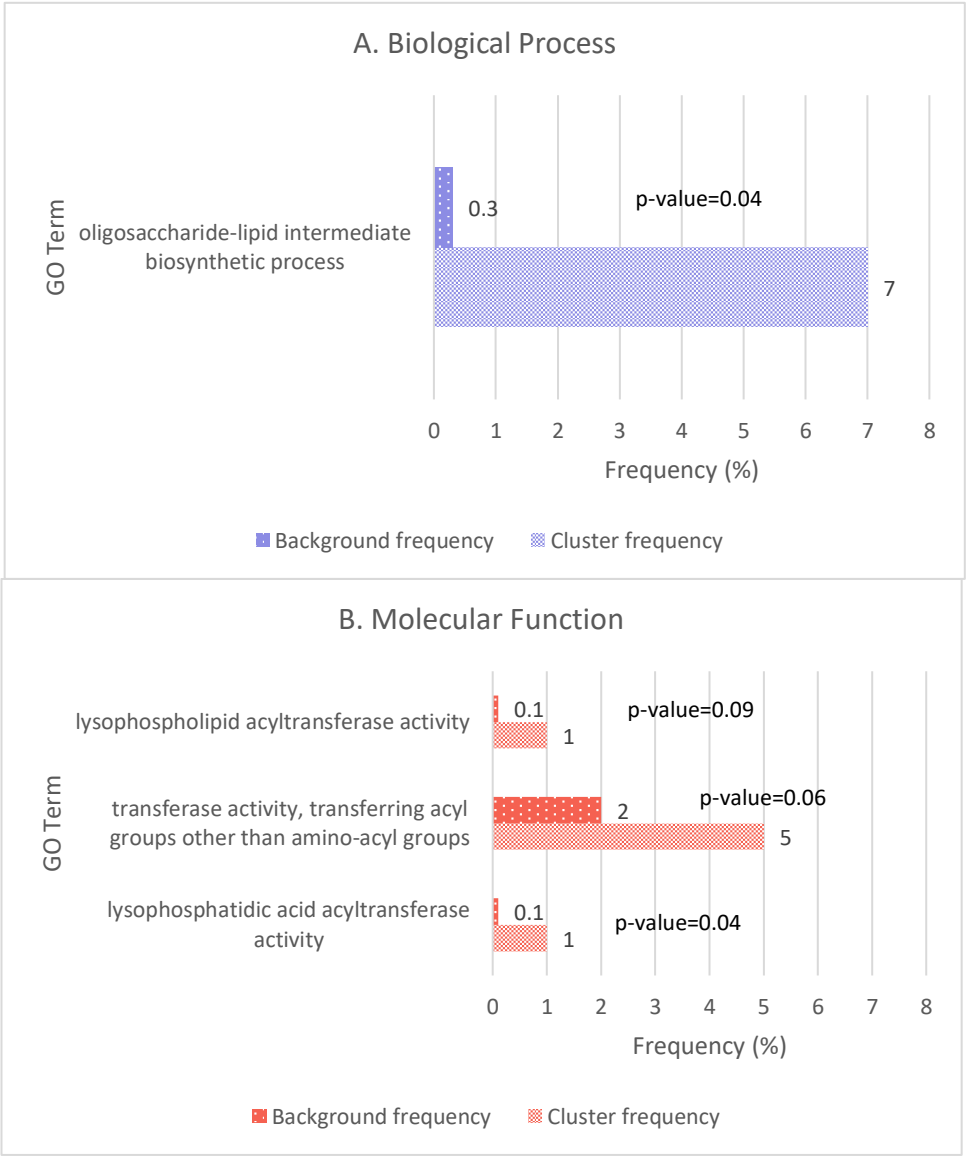


Figure 13. GO Term cluster frequencies of genes with unique missense SNPs, according to sample origin. A. GO Terms referent to Biological Process in samples obtained from the oral cavity; B. GO Terms referent to Molecular Function in samples obtained from vaginal exudates.

4. Antifungal susceptibility profiles

4.1. Determination of minimum inhibitory concentrations

The MICs for the isolates in study determined by Etest ranged from ≤0.002 to 0.38 µg/ml for AP, ≤0.002 to 0.004 µg/ml for AND, 0.002 to 0.032 µg/ml for MYC, 0.016 to 0.125 µg/ml for CS (Table 12), and 0.75 to ≥256 µg/ml for FL (Table 13).

Table 12. MICs (µg/ml) of antifungal drugs. Interpretation of results is also displayed: S for susceptible isolates; R for resistant isolates.

Isolate	Amphotericin B (AP) (µg/ml)	Caspofungin (CS) (µg/ml)	Anidulafungin (AND) (µg/ml)	Micafungin (MYC) (µg/ml)
5718fs	0.19 (S)	0.047 (S)	≤0.002 (S)	0.032 (S)
S145	0.125 (S)	0.023 (S)	≤0.002 (S)	0.012 (S)
S005a	0.38 (S)	0.047 (S)	≤0.002 (S)	0.032 (S)
7333fs	0.125 (S)	0.047 (S)	≤0.002 (S)	0.008 (S)
S142	0.023 (S)	0.047 (S)	≤0.002 (S)	0.006 (S)
SER7	0.25 (S)	0.125 (S)	≤0.002 (S)	0.003 (S)
S038	0.19 (S)	0.047 (S)	≤0.002 (S)	0.006 (S)
SER20	0.25 (S)	0.047 (S)	≤0.002 (S)	0.002 (S)
81	0.25 (S)	0.016 (S)	≤0.002 (S)	0.023 (S)
7248fs	0.095 (S)	0.064 (S)	≤0.002 (S)	0.023 (S)
SC5314	0.012 (S)	0.047 (S)	≤0.002 (S)	0.016 (S)

Table 13. MIC (µg/ml) of fluconazole, obtained with both Etest and broth microdilution. Interpretation of results is also displayed: S for susceptible isolates; R for resistant isolates.

Isolate	Etest (µg/ml)	Broth microdilution (µg/ml)
5718fs	2 (S)	0.125 (S)
S145	1.5 (S)	0.125 (S)
S005a	2 (S)	0.25 (S)
7333fs	4 (S)	0.5 (S)
S142	4 (S)	0.25 (S)
SER7	32 (R)	1 (I)
S038	3 (S)	0.125 (S)
SER20	1 (S)	0.125 (S)
81	0.75 (S)	0.5 (S)
7248fs	≥256 (R)	0.125 (S)
SC5314	≥256 (R)	0.125 (S)
T0	≥256 (R)	<0.5 (S)
ATCC 22019	ND	0.25 (S)

ND: Not determined.

The endpoint of amphotericin B tests was determined at 100% inhibition of growth (Figure 14A). For echinocandins, MICs were considered as the visual point at 80% inhibition, due to the possibility of microcolonies growing inside the ellipse, and interpretation of MIC should consider this trailing (Figure 14C, 14D, and 14E)[85]. The fluconazole MIC determination was hindered due to the high incidence of trailing observed (Figure 14B).

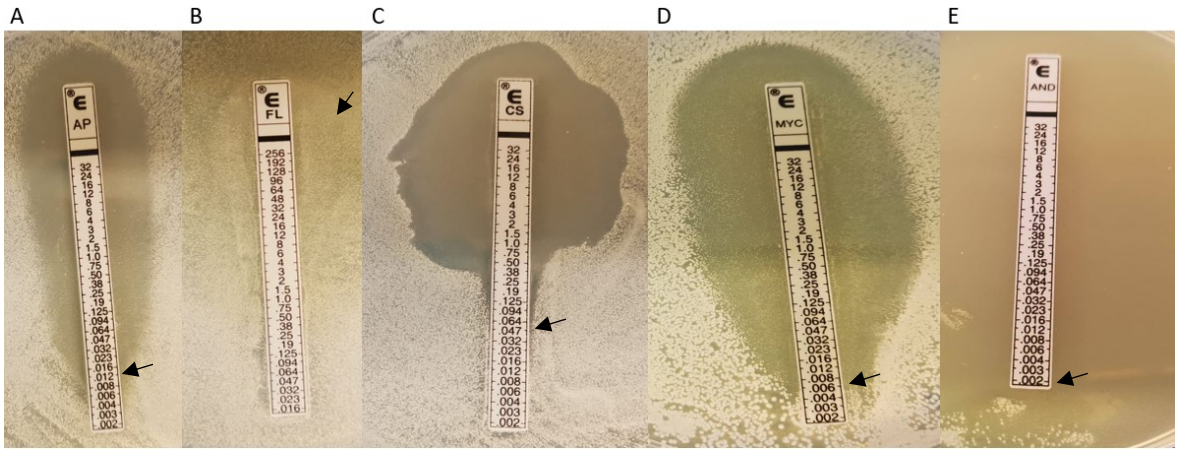


Figure 14. Etest results of some *Candida* isolates against the antifungals tested. (A) C. albicans control strain tested against amphotericin B. (B) *C. albicans* control strain tested against fluconazole. (C) *C. albicans* control strain tested against caspofungin. (D) *C. glabrata* isolate SER7 tested against anidulafungin. Arrows point to the endpoints considered, except in (B), where it points to the trailing region.

According to EUCAST guidelines[82] all isolates were susceptible to the antifungal drugs tested, except for fluconazole. The MICs of fluconazole for SER7 and 7248fs were above 16 µg/ml and 4 µg/ml, respectively, indicating they were resistant to this drug. The MICs for the *C. albicans* control strains SC5314 and strain T0 (both on RPMI and SD media) was ≥256 (Figure 15) which was not in agreement with the fluconazole sensitive strain *C. albicans* SC5314[86] with a reported MIC value of 0.24 µg/ml[87], and the reported MIC of 12 µg/ml for the T0 strain[32]. In this way, the MICs of fluconazole were determined by broth microdilution.

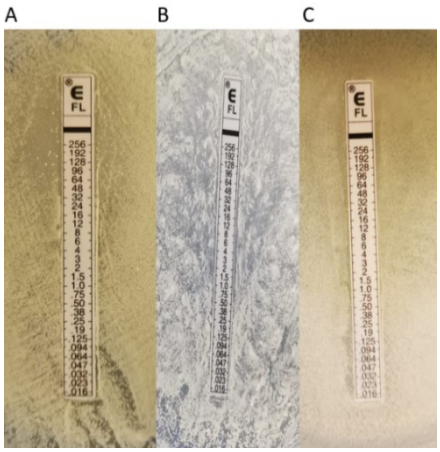


Figure 15. Etest results of *C. albicans* reference controls against fluconazole. (A) T0 on RPMI medium. (B) T0 on SD medium. (C) SC5314 on RPMI medium.

4.2. Broth microdilution assay for testing isolates' susceptibility to and determining MIC of fluconazole

As depicted in Table 13, the broth microdilution assays showed that all isolates were susceptible to fluconazole, with MICs ranging from 0.125 to 1 μg/ml. The MIC for strains SC5314 and T0 were 0.125 μg/ml and 0.5 μg/ml, respectively, which is within the range expected for these strains. Regarding the quality control strain *C. parapsilosis* ATCC 22019, the MIC was 0.25 μg/ml. Even though this value is outside the range expected[88], it is within ± one two-fold dilutions of the expected MIC range, and for that reason the results were considered acceptable[89].

.

IV. DISCUSSION

The genus *Candida* harbours species responsible for both superficial and systemic infections. The incidence of these infections has been increasing worldwide[90]. Despite *C. albicans* infections being more abundant[5], infections by non-*albicans Candida* species have increased, and special consideration has been given to infections caused by *C. glabrata*, since this species often shows higher tolerance to the antifungals more commonly used[91].

1. Whole-Genome Sequencing

Whole-genome sequence data and further analysis allow the identification of different types of mutations, such as single nucleotide polymorphisms, insertions, deletions, or structural variants, involved with either pathogenicity or drug resistance[92].

We performed whole-genome sequencing of ten *Candida albicans* clinical isolates to assess their genetic diversity. Sequencing was done with the MinION (Oxford Nanopore Technologies, UK), since it produces long reads and reliable detection of variants with fewer reads than those generally produced by other next-generation sequencing platforms[53].

Isolate SER7 was previously identified as *C. albicans* by phenotypic methods, but our NGS data identified it as being *C. glabrata*. Misidentification of *Candida* species is more common using conventional methods[93] with reports of atypical isolates of *Candida* being misidentified as *C. albicans* by conventional phenotypic tests[94,95]. In contrast, molecular identification methods have higher accuracy, sensitivity and specificity than phenotypic methods for the identification and differentiation of *C. albicans* from other *Candida* species[5]. A study by Bitar et. al (2014) aimed to identify and type *Candida albicans* isolates from two major hospitals in Beirut/Lebanon[96]. The authors reported a discrepancy in the identification of *Candida* isolates to the species level, using different methods: API (bioMérieux, France), CHROMagar cultures, and ITS sequencing, and comparing to the hospital identification carried out, which was based on the germ tube test, that differentiates between *C. albicans* and *Candida* non-*albicans*. Their results showed that samples tested on CHROMagar and germ tube test produced distinct identifications, isolates tested by API did not match the germ tube results; the samples were ITS sequenced for identification, and among *C. albicans* isolates, 21%, 15%, and 15% were misidentified by the germ tube test, by cultivation on CHROMagar, and by API, respectively. Some of these isolates identified as *C. albicans* were found to be *C. tropicalis, C. glabrata* or *C. sphearical*, and some isolates identified as *Candida* non-*albicans*

that were *C. albicans*. As supported by the authors, these type of results might lead to the use of inappropriate antifungal agents, which results in longer hospitalization and comorbidities[96].

Rapid and accurate identification of *Candida* species plays an important role in infection management of local or systemic candidiasis[5], and sequence-based identification techniques were found to be more efficient than other phenotypic methods[96]. In fact, identification to the species level allows a more accurate choice when applying the best antifungal therapy option for patients' successful recovery[97]. In the case of fluconazole, *C. glabrata* with MIC of 4 µg/ml would be categorised as "susceptible, increased exposure" whereas *C. albicans* would be categorised as "resistant"[82], which means that, for this example, the treatment with fluconazole would have no effect if the infection was caused by *C. albicans*, whereas if the infection was caused by *C. glabrata,* the treatment with fluconazole would have to be adjusted, since being "susceptible, increased exposure" means that there is a high likelihood of therapeutic success if the exposure to the antifungal is increased by adjusting the dosing regimen[69]. This reenforces the importance of accurate identification of *Candida* spp. for a successful management and treatment of infections, while this works suggests that ONT sequencing is appropriate for this purpose.

The MLST results did not allow us to assign a sequence type for *C. albicans* isolates and consequently we were not able to determine the clade to which the isolates belonged to; the relevant data was submitted to the *Candida albicans* PUBMLST database for the assignment of STs to the isolates. For *C. glabrata* isolate SER7, the MLST analysis revealed that this isolate belongs to ST-10, and further research in literature showed that ST 10 belongs to group 3[61]. European strains are more common in groups 1 and 3[64], however in group 3 an over-representation of isolates from the USA has been reported[61], although we had no information regarding the patient from whom the SER7 isolate was collected, namely if the patient was from Europe or the USA, or if the patient had recently travelled to or from Europe/USA. Additionally, there is no established association between antifungal resistance, particularly fluconazole resistance, and ST or group[61].

2. Single nucleotide polymorphisms

We found between 50320 and 104090 SNPs in the *Candida albicans* genomes studied, which is aligned with the 62000 – 70000 SNPs found in the genome of the *C. albicans* reference strain SC5314[98–100]. Considering that the *C. albicans* isolates studied were collected from different human body sites, and were certainly under different pressures from the host and the host environment, it can be expected that isolates have divergent numbers of SNPs, in particular of adaptive SNPs[100]. In the *Candida glabrata* isolate SER7 we found approximately 4.83 SNPs/kb, which was in accordance

to the range of 0.04 to 7.23 SNPs/kb SNPs reported for the genome of the *C. glabrata* reference strain CBS138[101].

In this study, we did not find SNPs previously associated with resistance to the antifungals tested[37–45]. This goes in line with the antifungal susceptibility profiles of the isolates, which were all susceptible to the antifungal drugs tested. It would be important to sequence also resistant strains, to validate our method as being able to call resistance-associated SNPs.

However, we found SNPs other than the ones previously described in resistance associated genes (*erg3*, resistance to echinocandins and azoles; *gsc1,* resistance to echinocandins; *tac1*, resistance to azoles). We do not know if these variants could play a role in resistance, for example through epistatic mechanisms[3]; but it could be interesting to conduct microevolution experiments focusing on these genes to evaluate if and how these variants may interact with other SNPs involved in antifungal resistance. Additionally, some of these SNPs could lead to future resistance by affecting the expression level of their associated genes, which could also be tested, either by blot analysis or RNA sequencing.

We explored the biological processes, molecular functions, and cellular components associated with the SNPs found in our strains.

There was a significantly higher frequency of genes with missense SNPs common to the three types of samples in the biological process of adhesion of a symbiont to its host. Adhesion to the host is one of *C. albicans* virulence factors[24]; *C. albicans* cells adhere to the host cell using adhesins, encoded by *als and hwp* genes[102]. In our isolates, the genes significantly involved in adhesion of a symbiont to host were *ace2, als1-4, hwp2* and *rfx2*. The gene *ace2* codes for a transcription factor involved in regulation of morphogenesis[18]; Kelly et al. (2004) reported that *ace2* null mutants have reduced ability to adhere to surfaces and altered cell wall biosynthesis, which affects adhesion and interaction with host cells[103]. The *als* genes encode a family of cell surface glycoproteins known as adhesins, many of which with a role in cell adhesion and adherence to host surfaces[18]; for example, *als1* codes for a cell surface protein that mediates adherence to endothelial and epithelial cells[104]; studies have shown that *als1* overexpression increases adherence by 125%[105]; *als3* encodes a cell wall adhesin which is closely related to Als1p in sequence, regulation and function[106]. The gene *hwp2* also codes for a cell wall protein[18] that has been reported as essential for adhesion to human epithelial and endothelial cells, and for biofilm formation[107]. *rfx2* encodes a RFX domain transcriptional repressor[18], and data has demonstrated that the product of *rfx2* plays important roles not only in the regulation of DNA damage responses, but also in morphogenesis, virulence, and

adherence, although the precise mechanisms were not fully understood[108]. However, we do not know whether our isolates were causing an infection or were commensals at the time of collection; the over representation of genes with missense SNPs involved in adhesion processes may justify the clinical significance of this group of isolates, since there is a strong correlation between adherence and ability to colonize and cause disease[102]. The fact that this group of genes shows an accumulated genetic variation in clinical strains might suggest that it is under selection to modulate its behaviour so that strains become more adapted to the host environment.

Regarding genes with missense SNPs unique to each isolate, according to their source, there was a significantly higher frequency of genes involved in oligosaccharide-lipid intermediate biosynthetic processes among isolates of the oral cavity. The genes *alg8* and *alg5* encode a putative glucosyltransferase involved in cell wall mannan biosynthesis[18], and the gene *ecm39* encodes a putative mannosyltransferase[18]. *C. albicans* biofilm matrix is composed by mannans, polysaccharides β-1,3-glucan and β-1,6-glucan, and chitin, forming the mannan-glucan complex (MGCx)[109,110]. In the oral cavity, hyphae formation and adherence to the oral epithelial cells promotes the development of biofilms, which once established promotes an increase in the expression of virulence factors, and a decrease in the antifungal susceptibility and phagocytosis[109], thus enhancing the fitness of the yeast to the oral cavity environment. This finding suggests the oral cavity isolates in study were well adapted to the niche they were collected from and had a different ability to form biofilms, which is important for *C. albicans* survival and pathogenicity[111].

Genes with missense SNPs uniquely found in *C. albicans* from vaginal exudates, whose Molecular Function was significantly enriched among our samples, were involved in lysophosphatidic acid acyltransferase activity. The gene *lpt1* encodes a lysophospholipid acyltransferase involved in phospholipid remodeling with a role in glycerophospholipid biosynthesis[18]. The product of *lpt1* has a major role in the lysophospholipid esterification, which in turn is associated with scavenging lysophospholipids produced by intracellular phospholipases, and utilizing lysophospholipids produced by extracellular phospholipases[112]. The latter is part of the hydrolytic enzymes that contribute to *C. albicans* virulence by having a role in host-cell membrane damage and in the adhesion of yeasts to host tissues[20,113], and in enhancing the efficiency of extracellular nutrient acquisition[21]. Lysophospholipase has also been associated with the formation of buds in rapidly growing cultures[114,115]. A study by Wong et. al (2019) concluded that *lpt1* deletion on *C. albicans* has an impact on its virulence, hence reinforcing the role of phospholipid remodeling and phospholipase products scavenging[113]. Since *C. albicans* mode of growth, i.e., yeast versus hypha, or biofilm versus planktonic, significantly changes the phospholipid profiles[112], the presence of SNPs in genes that

affect phospholipid remodeling might translate into a change in the *Candida* cells morphology, which could be indicative of a response by the isolates to improve colonization and invasion of host tissues.

In summary, we found SNPs in genes that were exclusive for vaginal exudate samples, as well as for samples originated from the oral cavity. The frequency of genes with the missense SNPs involved in molecular processes and biological functions was significantly superior (p-value ≤ 0.05) to the background frequency in those GO sets; the genes in question and their products were involved in mechanisms that might improve the isolates adaptability and survival ability in the environment they were collected from.

3. Antifungal susceptibility

All isolates were susceptible to all antifungals tested. A twelve-year study evaluated the prevalence of amphotericin B resistance of *Candida* species during 2006-2011 and 2012-2017, in Kuwait, and reported that the prevalence of resistance to amphotericin B during this period did not change and was very low – 0% for *C. albicans* and 1% for *C. glabrata*[116]. *C. glabrata* is usually susceptible to amphotericin B and tends to have higher MIC values than *C. albicans*[117], which was observed in 7 of the *C. albicans* isolates in our study.

The results for echinocandins susceptibility are in accordance with the prevalence of resistant *Candida* isolates described in the literature. Al-Baqsami and colleagues reported in their study that 93% of their *C. glabrata* isolates were susceptible to micafungin[118]. Another study pointed that the prevalence of *C. glabrata* resistance to anidulafungin, micafungin and caspofungin was 2%, 2% and 4% respectively; for *C. albicans* the prevalence of resistant isolates was 0% for anidulafungin, and 0.1% for micafungin and caspofungin[119,120].

There was a difficulty in reading Etest results when testing for fluconazole susceptibility, not only due to the observed trailing, but also due to unexpected results when testing reference controls, even when using different media and Etest strips from different batches. Previously, the impact of the heavy trailing effect in the interpretation of Etest results has also been seen by Song and colleagues for *Candida* isolates from patients with oral candidiasis[121]. To overcome this obstacle, we performed broth microdilution, which is the gold standard methodology for antifungal susceptibility testing[68], and allowed us to get interpretable results. All *C. albicans* isolates were susceptible to fluconazole, while *C. glabrata* was in the category "susceptible, increased exposure" which means that there is a high possibility of therapeutic success due to an increase of the fluconazole concentrations at the site of infection or by adjusting the dosing administration[69]. Despite this, in clinical environments, Etest or other gradient diffusion gadgets gain advantage since these are

inexpensive compared to broth microdilution, good for resource-limited settings and provide a MIC value, although the interpretation of its results is somewhat subjective. However, if the difficulties we have encountered were to happen in clinical settings, broth microdilution would be a more challenging technique to perform as plates preparation is labour-intensive and commercially prepared plates are very expensive, and technical training requirement is also significant high[68].

The incidence of resistance to fluconazole in the USA was between 0.5% to 2% for *C. albicans*, and between 11% to 13% for *C. glabrata*[122]. Between 2005 and 2007, global surveillance showed that 98% and 68% of *C. albicans* and *C. glabrata* isolates, respectively, were susceptible to fluconazole[123,124]. In their twelve-year study, Khan et al. (2019) reported that the prevalence of resistance to fluconazole during 2006-2011 and 2012-2017 increased from 3% to 5% for *C. albicans* and 3% to 14% for *C. glabrata*[116]. None of the isolates studied here was resistant to fluconazole, although the prevalence of *Candida* isolates resistant to fluconazole is rising worldwide, compromising the use of this drug which is the most used antifungal for the treatment of candidiasis[119].

V. CONCLUSION AND FUTURE WORK

The goal of this project was to study the genetic diversity of ten clinical fungal isolates identified as *Candida albicans*, using a comparative genomics approach, and to establish the antifungal susceptibility profiles of the isolates against different classes of antifungals.

All isolates were susceptible to the antifungal agents tested; also, the isolates did not have any of the polymorphisms described in the literature to be associated with antifungal resistance, which is in accordance with the antifungal susceptible profile traced, however, some SNPs were detected in genes that are associated to drug resistance. Further work on the effect of the identified SNPs on gene expression would be most interesting to clarify possible expression quantitative trait loci (eQTL). It would be interesting to differentiate between polymorphisms and causal mutations, since several mechanisms can lead to resistance and interact with each other - epistasis[125].

Whole genome sequencing and further analysis lead us to discover that one of the isolates tested was not a *C. albicans*, but instead a *C. glabrata*. Misidentification of *Candida* species may compromise optimal antifungal therapy, in a clinical setting, or influence how an experiment is performed and its results analysed, research wise. DNA sequencing methods are, most likely, one of the best options to identify *Candida* isolates to the species level. The methodology used (ONT), proved to be appropriate to achieve this, outperforming the classical methods.

It would be most interesting to increase the number of isolates from the different human sites, and to have information about the infection status of the hosts, to clarify whether the SNPs identified are indeed associated to the ecology of each particular niche (oral cavity, vaginal exudates and haemoculture) and to the ability to cause infection in humans; this could help identify genes that are relevant for the colonization and infection of each human site.

VI. REFERENCES

1. Turner SA, Butler G. The Candida pathogenic species complex. *Cold Spring Harb Perspect Med*. 2014;4(9):1-18. doi:10.1101/cshperspect.a019778
2. Brunke S, Hube B. Two unlike cousins: Candida albicans and C.glabrata infection strategies. *Cell Microbiol*. 2013;15(5):701-708. doi:10.1111/cmi.12091
3. Pais P, Galocha M, Viana R, Cavalheiro M, Pereira D, Teixeira MC. Microevolution of the pathogenic yeasts Candida albicans and Candida glabrata during antifungal therapy and host infection. *Microb Cell*. 2019;6(3):142-159. doi:10.15698/mic2019.03.670
4. Muñoz JF, Gade L, Chow NA, et al. Genomic insights into multidrug-resistance, mating and virulence in Candida auris and related emerging species. *Nat Commun*. 2018;9(1):1-13. doi:10.1038/s41467-018-07779-6
5. Neppelenbroek K, Seó R, Urban V, et al. Identification of Candida species in the clinical laboratory: a review of conventional, commercial, and molecular techniques. *Oral Dis*. 2013;20(4). doi:https://doi.org/10.1111/odi.12123
6. Mourad A, Perfect JR. What Can the Clinical Mycology Laboratory Do for Clinicians Today and Tomorrow? *Curr Clin Microbiol Reports*. 2017;4:96-105. doi:https://doi.org/10.1007/s40588-017-0061-y
7. Hadfield TL. *Medical Microbiology 18th Edition*. Vol 155.; 1990. doi:10.1093/milmed/155.7.a26
8. Barbosa MR. Evolution of antifungal drug resistance in Candida albicans. 2014. ria.ua.pt.
9. Kim J, Sudbery P. Candida albicans, a major human fungal pathogen. *J Microbiol*. 2011;49(2):171-177. doi:10.1007/s12275-011-1064-7
10. Sitterlé E, Maufrais C, Sertour N, Palayret M, d'Enfert C, Bougnoux ME. Within-Host Genomic Diversity of Candida albicans in Healthy Carriers. *Sci Rep*. 2019;9(1):1-12. doi:10.1038/s41598-019-38768-4
11. Odds FC, Brown AJ, Gow NA. Candida albicans genome sequence: a platform for genomics in the absence of genetics. *Genome Biol*. 2004;5(230). doi:https://doi.org/10.1186/gb-2004-5-7-230
12. Soll DR, Lockhart SR, Zhao R. Relationship between switching and mating in Candida albicans. *Eukaryot Cell*. 2003;2(3):390-397. doi:10.1128/EC.2.3.390-397.2003
13. Butler G, Rasmussen MD, Lin MF, et al. Evolution of pathogenicity and sexual reproduction in eight Candida genomes. *Nature*. 2010;459(7247):657-662. doi:10.1038/nature08064.Evolution
14. Usher J. The Mechanisms of Mating in Pathogenic Fungi-A Plastic Trait. *Genes (Basel)*. 2019;10(10):16-18. doi:10.3390/genes10100831
15. Noble SM, Johnson AD. Genetics of Candida albicans , a Diploid Human Fungal Pathogen . *Annu Rev Genet*. 2007;41(1):193-211. doi:10.1146/annurev.genet.41.042007.170146
16. Molero G, Dfcz-Orejas R, Navarro-Garcîa F, et al. Candida albicans: Genetics, dimorphism and pathogenicity. *Int Microbiol*. 1998;1(2):95-106. doi:10.2436/im.v1i2.54
17. Fidel Jr PL, A. VJ, Sobel JD. Candida glabrata: review of epidemiology, pathogenesis, and clinical disease with comparison to C. albicans. *Clin Microbiol Rev*. 1992;12(1):80-96.
18. Skrzypek M, Binkley J, Binkley G, Miyasato S, Simison M, Sherlock G. The Candida Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Res*. 2017;41. doi:10.1093/nar/gkw924
19. Carreté L, Ksiezopolska E, Gómez-Molero E, et al. Genome comparisons of Candida glabrata Serial Clinical Isolates Reveal Patterns of Genetic Variation in Infecting Clonal Populations. *Front Microbiol*. 2019. doi:https://doi.org/10.3389/fmicb.2019.00112

20. Gonçalves B, Ferreira C, Alves CT, Henriques M, Azeredo J, Silva S. Vulvovaginal candidiasis: Epidemiology, microbiology and risk factors. *Crit Rev Microbiol*. 2016;42(6):905-927. doi:10.3109/1040841X.2015.1091805

21. Mayer FL, Wilson D, Hube B. Candida albicans pathogenicity mechanisms. *Virulence*. 2013;4(2):119-128.

22. Pappas PG, Kauffman CA, Andes DR, et al. Clinical Practice Guideline for the Management of Candidiasis: 2016 Update by the Infectious Diseases Society of America. *Clin Infect Dis*. 2015;62(4):e1-e50. doi:10.1093/cid/civ933

23. Fidel Jr PL, Vazquez JA, Sobel JD. Candida glabrata: review of epidemiology, pathogenesis, and clinical disease with comparison to C. albicans. *Clin Microbiol Rev*. 1999;12(1):80-96.

24. Nikou SA, Kichik N, Brown R, et al. Candida albicans interactions with mucosal surfaces during health and disease. *Pathogens*. 2019;8(2):1-23. doi:10.3390/pathogens8020053

25. Rodrigues CF, Rodrigues ME, Silva S, Henriques M. Candida glabrata Biofilms: How Far Have We Come? *J Fungi*. 2017;3(1):11. doi:10.3390/jof3010011

26. Alves R, Kastora SL, Gomes-Gonçalves A, et al. Transcriptional responses to Candida glabrata biofilm cells to fluconazole are modulates by the carbon source. *npj Biofilms Microbiomes*. 2020;6(4). doi:https://doi.org/10.1038/s41522-020-0114-5

27. Brand A, Shanks S, Duncan VMS, Yang M, Mackenzie K, Gow NAR. Hyphal Orientation of Candida albicans Is Regulated by a Calcium-Dependent Mechanism. *Curr Biol*. 2007;17(4):347-352. doi:10.1016/j.cub.2006.12.043

28. Lachke SA, Joly S, Daniels K, Soll DR. Phenotypic switching and filamentation in Candida glabrata. *Microbiol*. 2002;148(9):2661-2674. doi:10.1099/00221287-148-9-2661

29. Pristov KE, Ghannoum MA. Resistance of Candida to azoles and echinocandins worldwide. *Clin Microbiol Infect*. 2019;25(7):792-798. doi:10.1016/j.cmi.2019.03.028

30. Renau TE, Hecker SJ, Lee VJ. Antimicrobial Potentiation Approaches: Targets and Inhibitors. In: *Annual Reports in Medicinal Chemistry Volume 33*. Vol 33. Elsevier Masson SAS; 1998:121-130. doi:10.1016/S0065-7743(08)61077-5

31. Berkow EL, Lockhart SR. Infection and Drug Resistance Dovepress Fluconazole resistance in Candida species: a current perspective. *Infect Drug Resist*. 2017;10(1):237-245. doi:10.2147/IDR.S118892

32. Weil T, Santamaría R, Lee W, et al. Adaptive Mistranslation Accelerates the Evolution of Fluconazole Resistance and Induces Major Genomic and Gene Expression Alterations in Candida albicans. *Msph Ther Prev*. 2017;2(4):1-14. doi:https://doi.org/10.1128/mSphere.00167-17

33. Anderson JB. Evolution of antifungal-drug resistance: Mechanisms and pathogen fitness. *Nat Rev Microbiol*. 2005;3(7):547-556. doi:10.1038/nrmicro1179

34. Campoy S, Adrio JL. Antifungals. *Biochem Pharmacol*. 2017;133:86-96. doi:10.1016/j.bcp.2016.11.019

35. Selmecki A, Forche A, Berman J. Aneuploidy and Isochromosome Formation in Drug-Resistant Candida albicans. *Science (80-)*. 2006;313(5785):367-370.

36. Bhattacharya S, Sae-Tia S, Fries BC. Candidiasis and Mechanisms ofAntifungal Resistance. *Antibiotics*. 2020;9(312). doi:10.3390/antibiotics9060312

37. Favre B, Didmon M, Ryder NS. Multiple amino acid substitutions in lanosterol 14α-demethylase contribute to azole resistance in Candida albicans. *Microbiology*. 1999;145(10):2715-2725. doi:10.1099/00221287-145-10-2715

38. Spettel K, Barousch W, Makristathis A, et al. Analysis of antifungal resistance genes in Candida albicans and Candida glabrata using next generation sequencing. *PLoS One*. 2019;14(1):1-19. doi:10.1371/journal.pone.0210397

39. Manastir L, Ergon MC, Yucesoy M. Investigation of mutations in Erg11 gene of fluconazole resistant Candida albicans isolates from Turkish hospitals. *Mycoses*. 2011;54(2). doi:https://doi.org/10.1111/j.1439-0507.2009.01766.x

40. Flowers SA, Colón B, Whaley SG, Schuler MA, Rogers PD. Contribution of Clinically Derived Mutations in ERG11 to Azole Resistance in Candida albicans. *Antimicrob Agents Chemother*. 2015;59:450-460. doi:10.1128/AAC.03470-14.

41. Morio F, Pagniez F, Lacroix C, Miegeville M, Le Pape P. Amino acid substitutions in the Candida albicans sterol Δ5,6-desaturase (Erg3p) confer azole resistance: characterization of two novel mutants with impaired virulence. *J Antimicrob Chemother*. 2012;67(9):2131-2138. doi:https://doi.org/10.1093/jac/dks186

42. Coste AT, Turner V, Ischer F, et al. A Mutation in Tac1p, a Transcription Factor Regulating CDR1 and CDR2, Is Coupled With Loss of Heterozygosity at Chromosome 5 to Mediate Antifungal Resistance in Candida albicans. *Genetics*. 2006;172(4):2139-2156. doi:10.1534/genetics.105.054767

43. Siikala E, Rautemaa R, Richardson M, Saxen H, Bowyer P, Sanglard D. Persistent Candida albicans colonization and molecular mechanisms of azole resistance in autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy (APECED) patients. *J Antimicrob Chemother*. 2010;65(12):2505-2513. doi:https://doi.org/10.1093/jac/dkq354

44. Coste AT, Crittin J, Bauser C, Rohde B, Sanglard D. Functional Analysis of cis- and trans-Acting Elements of the Candida albicans CDR2 Promoter with a Novel Promoter Reporter System. *Eukaryot Cell*. 2009. doi:https://doi.org/10.1128/EC.00069-09

45. Ferrari S, Ischer F, Calabrese D, et al. Gain of Function Mutations in CgPDR1 of Candida glabrata Not Only Mediate Antifungal Resistance but Also Enhance Virulence. *PLoS Pathog*. 2009. doi:https://doi.org/10.1371/journal.ppat.1000268

46. Ellis D. Amphotericin B: spectrum and resistance. *J Antimicrob Chemother*. 2002;49(suppl_1):7-10. doi:10.1093/jac/49.suppl_1.7

47. Wiederhold NP, Najvar LK, Bocanegra RA, Kirkpatrick WR, Patterson TF. Caspofungin Dose Escalation for Invasive Candidiasis Due to Resistant Candida albicans. *Antimicrob Agents Chemother*. 2011;55(7):3254-3260. doi:10.1128/AAC.01750-10

48. Garnaud C, Botterel F, Sertour N, et al. Next-generation sequencing offers new insights into the resistance of Candida spp. to echinocandins and azoles. *J Antimicrob Chemother*. 2015;70(9):2556-2565. doi:https://doi.org/10.1093/jac/dkv139

49. Dudiuk C, Gamarra S, Jimenez-Ortigosa C, et al. Quick detection of FKS1 mutations responsible for clinical echinocandin resistance in Candida albicans. *J Clin Microbiol*. 2015;52:2037-2041. doi:10.1128/JCM.00398-15

50. Spreghini E, Orlando F, Sanguinetti M, et al. Comparative Effects of Micafungin, Caspofungin, and Anidulafungin against a Difficult-To-Treat Fungal Opportunistic Pathogen, Candida glabrata. *Antimicrob Agents Chemother*. 2012;56(3):1215-1222. doi:10.1128/AAC.05872-11

51. Fricker AM, Podlesny D, Fricke WF. What is new and relevant for sequencing-based microbiome research? A mini-review. *J Adv Res*. 2019;19:105-112. doi:10.1016/j.jare.2019.03.006

52. Schurch AC, Arredondo-Alonso S, Willems RJL, Goering R V. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene approaches. *Clin Microbiol Infect*. 2018;24(4):350-354. doi:10.1016/j.cmi.2017.12.016

53. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol*. 2016:1-11. doi:10.1186/S13059-016-1103-0

54. Tan S, Dvorak CMT, Estrada AA, Gebhart C, Marthaler DG, Murtaugh MP. MinION sequencing of Streptococcus suis allows for functional characterization of bacteria by multilocus sequence typing and antimicrobial resistance profiling. *J Microbiol Methods*. 2020;169(November 2019):105817.

doi:10.1016/j.mimet.2019.105817

55. Panthee S, Hamamoto H, Ishijima SA, Paudel A, Sekimizu K. Utilization of Hybrid Assembly Approach to Determine the Genome of an Opportunistic Pathogenic Fungus, Candida albicans TIMM 1768. *Genome Biol Evol*. 2018;10(8):2017-2022. doi:10.1093/gbe/evy166

56. Chandak S, Neu J, Tatwawadi K, et al. Overcoming High Nanopore Basecaller Error Rates for DNA Storage Via Basecaller-Decoder Integration and Convolutional Codes. *bioRxiv*. 2020. doi:https://doi.org/10.1101/2019.12.20.871939

57. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol*. 2019;20(129). https://doi.org/10.1186/s13059-019-1727-y.

58. Bougnoux ME, Tavanti A, Bouchier C, et al. Collaborative Consensus for Optimized Multilocus Sequence Typing of Candida albicans. *J Clin Microbiol*. 2003;41(11):5265-5266. doi:10.1128/JCM.41.11.5265-5266.2003

59. PubMLST Public databases for molecular typing and mocrobial genome diversity: Multi-Locus Sequence Typing. https://pubmlst.org/multilocus-sequence-typing.

60. Odds FC, Jacobsen MD. Multilocus sequence typing of pathogenic Candida species. *Eukaryot Cell*. 2008;7(7):1075-1084. doi:10.1128/EC.00062-08

61. Dodgson AR, Pujol C, Denning DW, Soll DR, Fox AJ. Multilocus sequence typing of Candida glabrata reveals geographically enriched clades. *J Clin Microbiol*. 2003;41(12):5709-5717. doi:10.1128/jcm.41.12.5709-5717.2003

62. Jolley KA, Bray JE, Maiden MCJ. PubMLST Public databases for molecular typing and microbial genome diversity. Wellcome Open Res. doi:https://doi.org/10.12688/wellcomeopenres.14826.1

63. Odds FC. Molecular phylogenetics and epidemiology of Candida albicans. *Future Microbiol*. 2010:67-79.

64. Dodgson AR, Pujol C, Pfaller MA, Denning DW, Soll DR. Evidence for recombination in Candida glabrata. *Fungal Genet Biol*. 2005;42:233-243. doi:doi:10.1016/j.fgb.2004.11.010

65. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: a tool for the unification of biology. *Nat Genet*. 2000;25(1):25-29. doi:10.1038/75556

66. Ashburner M, Ball CA, Blake JA, et al. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res*. 2021;49:325-334. doi:10.1093/nar/gkaa1113

67. Inglis D, Skrzypek M, Arnaud M, et al. Improved Gene Ontology Annotation for Biofilm Formation, Filamentous Growth, and Phenotypic Switching in Candida albicans. *Eukaryot Cell*. 2012;12(1):101-108. doi:10.1128/EC.00238-12

68. Berkow EL, Lockhart SR, Ostrosky-Zeichner L. Antifungal susceptibility testing: Current approaches. *Clin Microbiol Rev*. 2020;33(3):1-30. doi:10.1128/CMR.00069-19

69. Subcommittee on Antifungal Susceptibility Testing (AFST) of the ESCMID European Committee for Antimicrobial Susceptibility Testing. EUCAST antifungal MIC method for yeasts: Method for the determination of broth dilution minimum inhibitory concentrations of antifungal agents for yeasts. *EUCAST*. 2020;(April):1-21.

70. Alastruey-Izquierdo A, Melhem MSC, Bonfietti LX, Rodriguez-Tudela JL. SUSCEPTIBILITY TEST FOR FUNGI: CLINICAL AND LABORATORIAL CORRELATIONS IN MEDICAL MYCOLOGY. *Rev Inst Med Trop Sao Paulo*. 2015;57(1):57-64. doi:10.1590/S0036-46652015000700011

71. NCCLS. Reference Method for Broth Dilution Antifungal Susceptibility Testing of Yeasts ; Approved Standard. *NCCLS Doc M27-A2*. 2002;22(15).

72. De Coster W, D'Hert S, Schultz DT, Cruts M, M & Van Broeckhoven C. Nanopack: Visualizing and processing long-read sequencing data. *Bioinformatics*. 2018;34(15):2666-2669. doi:https://doi.org/10.1093/bioinformatics/bty149

73. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via k-mer weighting and repeat separation. *Genome Res*. 2017;27(5). doi:10.1101/gr.215087.116

74. Kolmogorov M, Yuan J, Lin Y, Pevzner P. Assembly of Long Error-Prone Reads Using Repeat Graphs. *Nat Biotechnol*. 2019. doi:doi:10.1038/s41587-019-0072-8

75. Vaser R, Nagarajan N, Sikic M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*. 2017;(27):737-746. doi:doi:10.1101/gr.214270.116

76. Medaka. ttps://github.com/nanoporetech/medaka. Published 2018.

77. Walker BJ, Abeel T, Shea T, et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS One*. 2014. doi:https://doi.org/10.1371/journal.pone.0112963

78. Li H. Minimap2: fast pairwise alignment for long nucleotide sequences. *Bioinformatics*. 2018;34:3094-3100. doi:10.1093/bioinformatics/bty191

79. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10(421). doi:https://doi.org/10.1186/1471-2105-10-421

80. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res*. 2018;3(124). doi:10.12688/wellcomeopenres.14826.1

81. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff SNPs in the genome of Drosophila melanogaster strains w1118; iso-2; iso-3. *Fly*. 2012;6(2):80-92. doi:10.4161/fly.19695

82. The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs for antifungal agents, version 10.0, 2020. *Eucast*. 2020;(April):0-8. http://www.eucast.org/astoffungi/clinicalbreakpointsforantifungals.

83. Bezerra AR, Simões J, Lee W, et al. Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification. *Proc Natl Acad Sci U S A*. 2013;110(27):11079-11084. doi:10.1073/pnas.1302094110

84. Binns D, Dimmer E, Huntley R, Barrel D, O'Donovan C, Apweiler R. QuickGO: a web-based tool for Gene Ontology seraching. *Bioinformatics*. 2009;25(22):3045-3046. doi:10.1093/bioinformatics/btp536

85. AB BIODISK. Etest for MIC Determination of Antifungal Agents: Reading Guide for Yeast. 2006.

86. Bandara HMHN, Wood DLA, Vanwonterghem I, Hugenholtz P, Cheung BPK, Samaranayake LP. Fluconazole resistance in Candida albicans is induced by Pseudomonas aeruginosa quorum sensing. *Sci Rep*. 2020;10(1):1-17. doi:10.1038/s41598-020-64761-3

87. Calabrese EC, Castellano S, Santoriello M, et al. Antifungal activity of azole compounds CPA18 and CPA109 against azole-susceptible and -resistant strains of Candida albicans. *J Antimicrob Chemother*. 2013;68(5):1111-1119. doi:https://doi.org/10.1093/jac/dks506

88. The European Committee on Antimicrobial Susceptibility Testing. Routine and extended internal quality control for MIC determination and disk diffusion as recommended by EUCAST. 2020;Version 5. http://www.eucast.org.

89. Andrews JM. Determination of minimum inhibitory concentrations. *J Antimicrob Chemother*. 2001;48(SUPPL. 1):5-16. doi:10.1093/jac/48.suppl_1.5

90. Pristov KE, Ghannoum MA. Resistance of Candida to azoles and echinocandins worldwide. *Clin Microbiol Infect*. 2019;25(7):792-798. doi:10.1016/j.cmi.2019.03.028

91. Henriques M, Williams D. Pathogenesis and virulence of candida albicans and candida glabrata. *Pathogens*. 2020;9(9):1-3. doi:10.3390/pathogens9090752

92. Sanglard D. Finding the needle in a haystack: Mapping antifungal drug resistance in fungal pathogen by genomic approaches. *PLoS Pathog*. 2019;15(1):1-9. doi:10.1371/journal.ppat.1007478

93. Huang Y-S, Wang F-D, Chen YC, et al. High rates of misidentification of uncommon Candida species causing bloodstream infections using conventional phenotypic methods. *J Formos Med Assoc*. 2020. https://doi.org/10.1016/j.jfma.2020.11.002.

94. Coleman DC, Sullivan DJ, Bennett DE, Moran GP, Barry HJ, Shanley DB. Candidiasis: the emergence of a novel species, Candida dubliniensis. *AIDS*. 1997;11(5):557-567.

95. McCullough M, Ross B, Reade P. Characterization of Genetically Distinct Subgroup of Candida albicans Strains Isolated from Oral Cavities of Patients Infected with Human Immunodeficiency Virus. *J Clin Microbiol*. 1995;33(3):696-700.

96. Bitar I, Khalaf RA, Harastani H, Tokajian S. Identification, Typing, Antifungal Resistance Profile, and Biofilm Formation of Candida albicans Isolates from Lebanese Hospital Patients. *Biomed Res Int*. 2014;2014:10. doi:Identification, Typing, Antifungal Resistance Profile, and Biofilm Formation of Candida albicans Isolates from Lebanese Hospital Patients

97. Lockhart SR, Jackson BR, Vallabhaneni S, Ostrosky-Zeichner L, Pappas PG, Chiller T. Thinking beyond the Common Candida Species: Need for Species-Level Identification of Candida Due to the Emergence of Multidrug-Resistant Candida auris. *J Clin Microbiol*. 2017;55(12):3324-3327. doi:10.1128/JCM.01355-17

98. Muzzey D, Schwartz K, Weissman JS, Sherlock G. Assembly of a phased diploid Candida albicans genome facilitates allele-specific measurements and provides a simple model for repeat and indel structure. *Genome Biol*. 2013;14(9). doi:10.1186/gb-2013-14-9-r97

99. Ciudad T, Hickman M, Bellido A, Berman J, Larriba G. Phenotypic Consequences of a Spontaneous Loss of Heterozygosity in a Common Laboratory Strain of Candida albicans. *Genetics*. 2016;203(3):1161-1176. doi:https://doi.org/10.1534/genetics.116.189274

100. Forche A, Magee PT, Magee BB, May G. Genome-Wide Single-Nucleotide Polymorphism Map for Candida albicans. *Eukaryot Cell*. 2004;3(3):705-714. doi:10.1128/EC.3.3.705-714.2004

101. Carreté L, Ksiezopolska E, Pegueroles C, et al. Patterns of Genomic Variation in the Opportunistic Pathogen Candida glabrata Suggest the Existence of Mating and a Secondary Association with Humans. *Curr Biol*. 2018;28(1):15-27. doi:https://doi.org/10.1016/j.cub.2017.11.027

102. Martin H, Kavanagh K, Velasco-Torrijos T. Targeting adhesion in fungal pathogen Candida albicans. *Future Med Chem*. 2020. doi:https://doi.org/10.4155/fmc-2020-0052

103. Kelly MT, MacCallum DM, Clancy SD, Odds FC, Brown AJP, Butler G. The Candida albicans CaACE2 gene affects morphogenesis, adherence and virulence. *Mol Microbiol*. 2004;53(3). doi:https://doi.org/10.1111/j.1365-2958.2004.04185.x

104. Sheppard DC, Yeaman MR, Welch WH, Zhang M, Waring AJ, Edwards Jr. JE. Functional and Structural Diversity in the Als Protein Family of Candida albicans. *Mol Basis Cell Dev Biol*. 2004;279(29):30480-30489. doi:https://doi.org/10.1074/jbc.M401929200

105. Fu Y, Ibrahim AS, Sheppard DC, et al. Candida albicans Als1p: an adhesin that is a downstream effector of the EFG1 filamentation pathway. *Mol Microbiol*. 2002;44(1):61-72. doi:10.1046/j.1365-2958.2002.02873.x.

106. Nobile CJ, Schneider HA, Nett JE, et al. Complementary adhesin function in C. albicans biofilm formation. *Curr Biol*. 2008;18(14):1017-1024. doi:10.1016/j.cub.2008.06.034.

107. Younes S, Bahnan W, Dimassi HI, Khalaf RA. The Candida albicans Hwp2 is necessary for proper adhesion, biofilm formation and oxidative stress tolerance. *Microbiol Res*. 2011;166(5):430-436. doi:10.1016/j.micres.2010.08.004

108. Hao B, Clancy CJ, Cheng S, Raman SB, Iczkowski KA, Nguyen MH. Candida albicans RFX2 Encodes a DNA Binding Protein Involved in DNA Damage Responses, Morphogenesis, and Virulence. *Eukaryot*

Cell. 2009;8(4):627-639. doi:10.1128/EC.00246-08

109. Vila T, Sultan AS, Montelongo-Jauregui D, Jabra-Rizk MA. Oral candidiasis: A disease of opportunity. *J Fungi*. 2020;6(1):1-28. doi:10.3390/jof6010015

110. Ene I V., Adya AK, Wehmeier S, et al. Host carbon sources modulate cell wall architecture, drug resistance and virulence in a fungal pathogen. *Cell Microbiol*. 2012;14(9):1319-1335. doi:10.1111/j.1462-5822.2012.01813.x

111. Gulati M, Nobile CJ. Candida albicans biofilms: development, regulation, and molecular mechanisms. *Microbes Infect*. 2016;18(5):310-321. doi:10.1016/j.micinf.2016.01.002

112. Ayyash M, Algahmi A, Gillespie J, Oelkers P. Characterization of a lysophospholipid acyltransferase involved in membrane remodeling in Candida albicans. *Biochim Biophys Acta*. 2014;1841(4):505-513. doi:10.1016/j.bbalip.2013.12.015

113. Wong D, Plumb J, Talab H, Kurdi M, Pokhrel K, Oelkers P. Genetically Compromising Phospholipid Metabolism Limits Candida albicans' Virulence. *Mycopathologia*. 2019;184:213-226. doi:https://doi.org/10.1007/s11046-019-00320-3

114. Pugh D, Cawson RA. The cytochemical localization of phospholipase in Candida albicans infecting the chick chorio-allantoic membrane. *Sabouraudia*. 1977;15:29-35.

115. Niewerth M, Korting HC. Phospholipases of Candida albicans. *Mycoses*. 2001;44:361-367. doi:doi:10.1046/j.1439-0507.2001.00685.x

116. Khan Z, Ahmad S, Al-Sweih N, et al. Changing trends in epidemiology and antifungal susceptibility patterns of six bloodstream Candida species isolates over a 12-year period in Kuwait. *PLoS One*. 2019;14(5):1-16. doi:10.1371/journal.pone.0216250

117. O'Shaughnessy EM, Lyman CA, Walsh TJ. Amphotericin B: Polyene Resistance Mechanisms. *Antimicrob Drug Resist*. 2009:295-305. doi:10.1007/978-1-59745-180-2_25

118. Al-Baqsami ZF, Ahmad S, Khan Z. Antifungal drug susceptibility, molecular basis of resistance to echinocandins and molecular epidemiology of fluconazole resistance among clinical Candida glabrata isolates in Kuwait. *Sci Rep*. 2020;10(1):5-7. doi:10.1038/s41598-020-63240-z

119. Pfaller MA, Diekema DJ, Turnidge JD, Castanheira M, Jones RN. Twenty years of the SENTRY Antifungal Surveillance Program: Results for Candida species from 1997-2016. *Open Forum Infect Dis*. 2019;6(Suppl 1):S79-S94. doi:10.1093/ofid/ofy358

120. Pfaller MA, Messer SA, Woosley LN, Jones RN, Castanheira M. Echinocandin and Triazole Antifungal Susceptibility Profiles for Clinical Opportunistic Yeast and Mold Isolates Collected from 2010 to 2011: Application of New CLSI Clinical Breakpoints and Epidemiological Cutoff Values for Characterization of Geographic . *J Clin Microbiol*. 2013;51(8):2571-2581. doi:10.1128/JCM.00308-13

121. Song YB, Suh MK, Ha GY, Kim H. Antifungal susceptibility testing with etest for Candida species isolated from patients with oral candidiasis. *Ann Dermatol*. 2015;27(6):715-720. doi:10.5021/ad.2015.27.6.715

122. Castanheira M, Deshpande LM, Davis AP, Rhomberg PR, Pfaller MA. Monitoring antifungal resistance in a global collection of invasive yeasts and molds: Application of CLSI epidemiological cutoff values and whole-genome sequencing analysis for detection of azole resistance in Candida albicans. *Antimicrob Agents Chemother*. 2017;61(10):1-20. doi:10.1128/AAC.00906-17

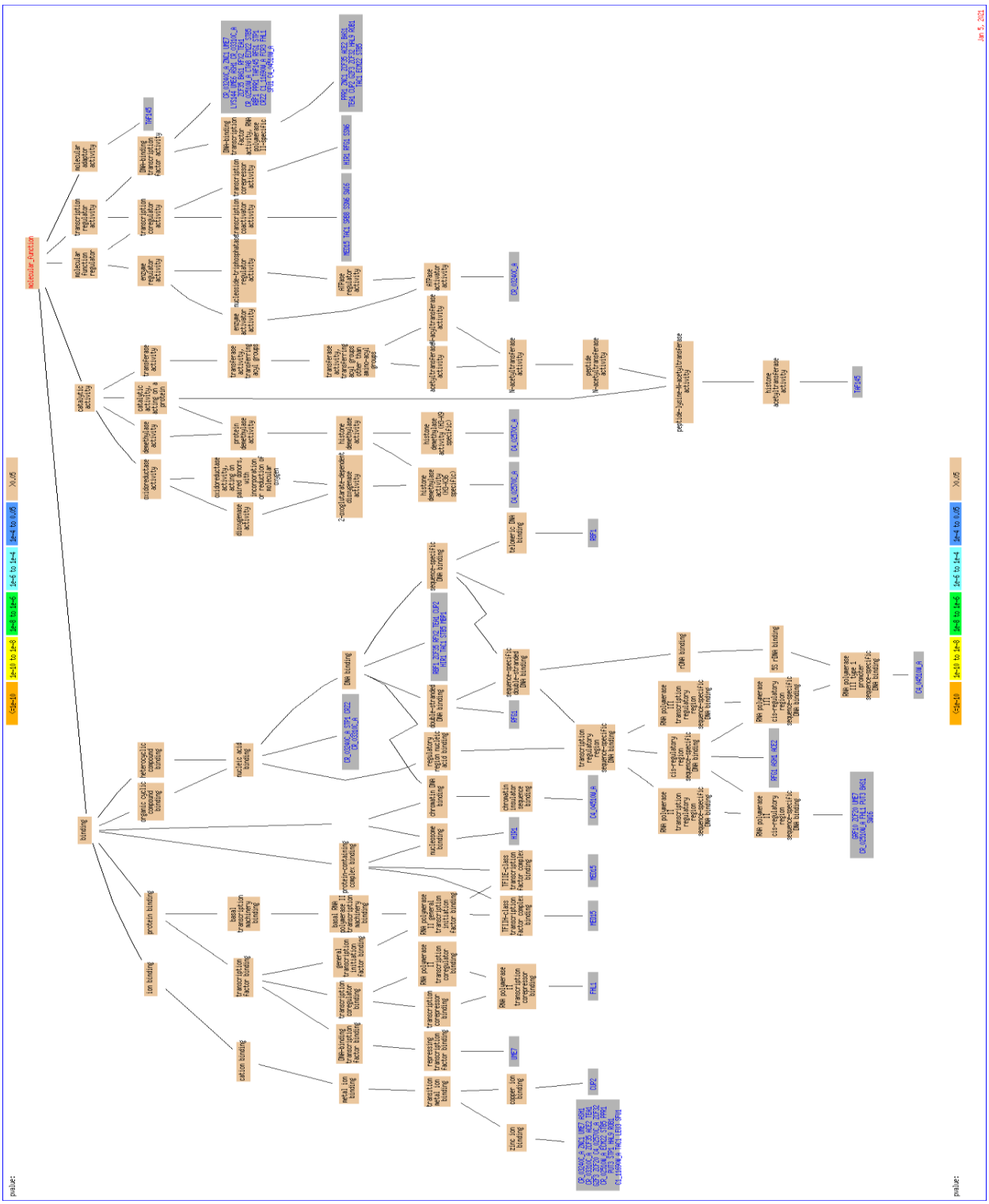
123. Beardsley J, Halliday CL, Chen SCA, Sorrell TC. Responding to the emergence of antifungal drug resistance: Perspectives from the bench and the bedside. *Future Microbiol*. 2018;13(10):1175-1191. doi:10.2217/fmb-2018-0059

124. Pfaller MA, Diekema DJ, Gibbs DL, et al. Results from the artemis disk global antifungal surveillance study, 1997 to 2007: A 10.5-year analysis of susceptibilities of candida species to fluconazole and voriconazole as determined by CLSI standardized disk diffusion. *J Clin Microbiol*. 2010;48(4):1366-1377. doi:10.1128/JCM.02117-09

125. Spettel K, Barousch W, Makristathis A, et al. Analysis of antifungal resistance genes in Candida albicans and Candida glabrata using next generation sequencing. *PLoS One*. 2019;14(1):1-19. doi:10.1371/journal.pone.0210397

VII. APPENDIX

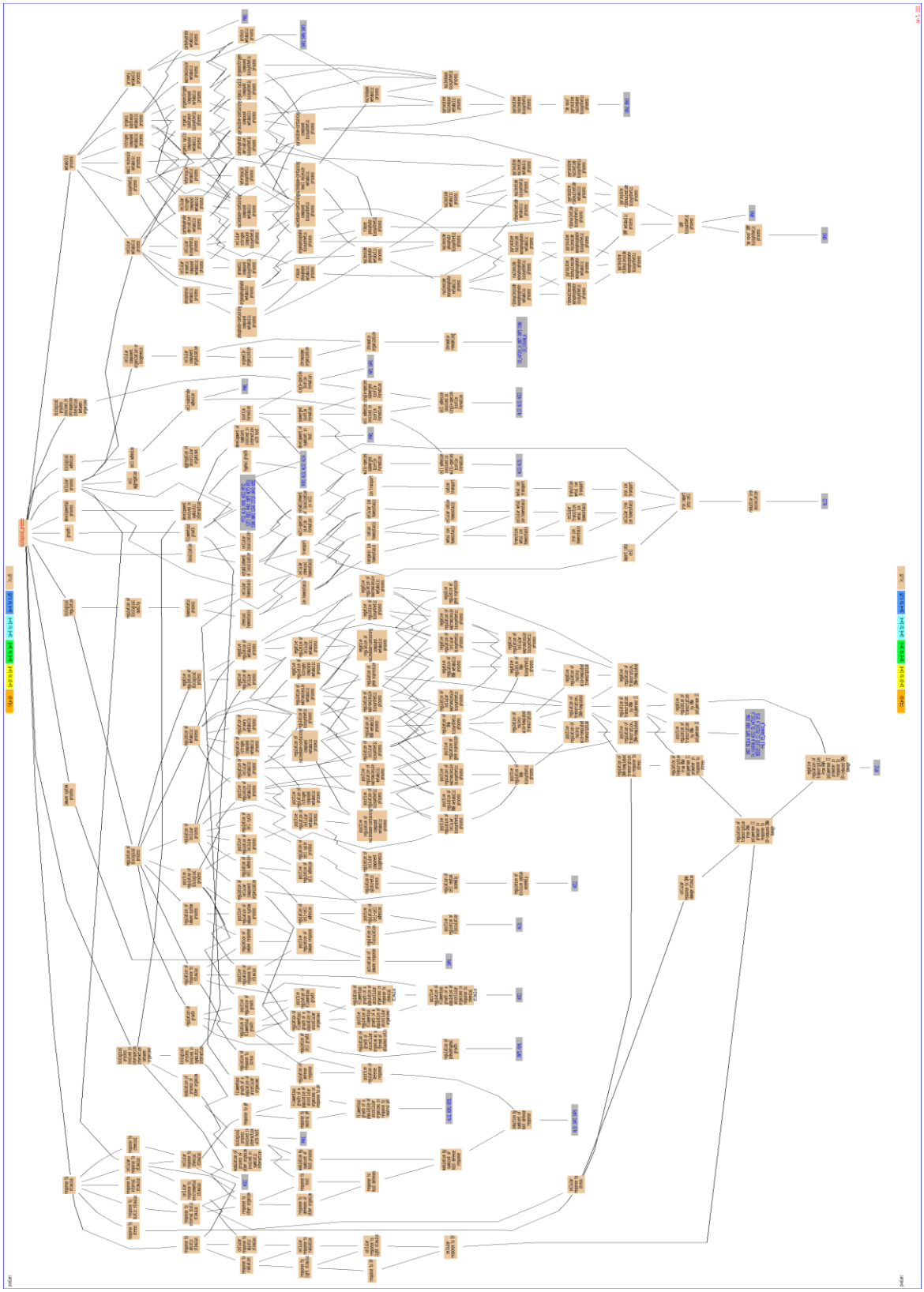
Appendix 1. Graphic displaying the relationships among the GO terms related to molecular function associated to genes with missense SNPs common to all samples. The analysis was done on January 5th, 2021.



Appendix 1.1. Information regarding genes with missense SNPs common to all samples, annotated to GO terms from the molecular function ontology.

GO term	Cluster frequency	Background frequency	P-value	Genes annotated to the term
Transcription regulator activity	47 out of 664 genes, 7.1%	269 out of 6473 background genes, 4.2%	0.06	*stb5, ume6, put3, taf145, med15, swi6, cta8, sfu1, c1_11690w_a, rob1, cup2, hir1, fhl1, ash1, lys144, mbp1, stp1, zcf32, tea1, zcf35, ssn6, rfx2, ume7, c4_02570c_a, c4_04510w_a, bas1, srb8, zcf20, fcr3, hal9, tac1, znc1, leu3, ifh1, med14, ppr1, grf10, rbf1, cr_02510w_a, rfg1, gzf3, cr_03240c_a, cr_03310c_a, med5, crz2, ace2, ecm22*
DNA-binding transcription factor activity	40 out of 664 genes, 6.0%	218 out of 6473 background genes, 3.4%	0.06	*stb5, ume6, put3, taf145, swi6, cta8, sfu1, c1_11690w_a, rob1, cup2, fhl1, ash1, lys144, mbp1, stp1, zcf32, tea1, zcf35, rfx2, ume7, c4_02570c_a, c4_04510w_a, bas1, zcf20, fcr3, hal9, tac1, znc1, leu3, ppr1, grf10, rbf1, cr_02510w_a, rfg1, gzf3, cr_03240c_a, cr_03310c_a, crz2, ace2, ecm22*

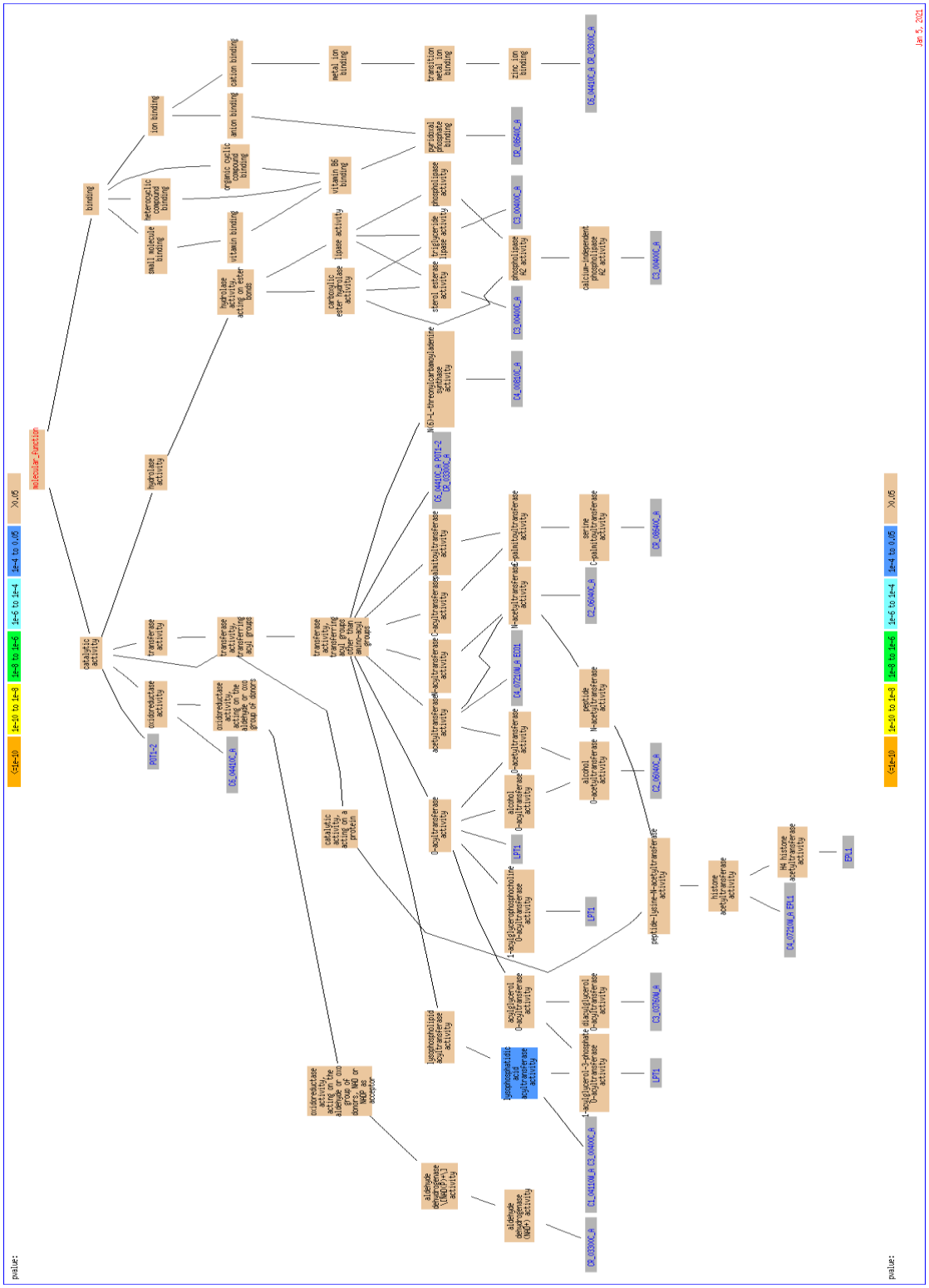
Appendix 1.2. Graphic displaying the relationships among the GO terms related to biological process associated to genes with missense SNPs common to all samples. The analysis was done on January 5th, 2021.



Appendix 1.2.1. Information regarding genes with missense SNPs common to all samples, annotated to GO terms from the biological process ontology.

GO term	Cluster frequency	Background frequency	P-value	Genes annotated to the term
Adhesion of symbiont to host	16 out of 664 genes, 2.4%	49 out of 6473 background genes, 0.8%	0.03	*ace2, als1, als2, als3, als4, his4, hwp2, pde1, phr1, rfx2, sap2, sap3, sap6, snf5, ura3, ywp1*
Negative regulation of macromolecule metabolic process	68 out of 664 genes, 10.2%	413 out of 6473 background genes, 3.4%	0.06	*c1_05660c_a, ire1, ccr4, med15, not5, ptc1, c1_09390w_a, c1_09790c_a, sfu1, gin1, c1_12370w_a, kem1, hir1, sas2, c2_05220c_a, fhl1, c2_07690w_a, sup35, gcn1, ash1, c3_01770c_a, rna1, rad50, ycg1, zcf32, c3_05320w_a, ssn6, c4_01660w_a, c4_01780c_a, rfx2, c4_02400c_a, ume7, c4_02570c_a, dot4, srb8, puf3, mph1, ago1, c4_06850c_a, ybl053, zcf20, orc4, leu3, cst20, med14, c5_04640c_a, ssd1, c6_00310w_a, ctf5, c6_02430w_a, nab3, c6_04060w_a, cr_02550c_a, rfg1, gzf3, cr_02930w_a, cr_03200c_a, yku80, cr_04720c_a, cr_06470w_a, cr_06960w_a, pif1, ace2, cr_07600w_a, asf1, ski2, cr_08930c_a, cr_09490w_a*
Negative regulation of nucleobase-containing compound metabolic process	48 out of 664 genes, 7.2%	266 out of 6473 background genes, 4.1%	0.09	*c1_05660c_a, med15, c1_09390w_a, sfu1, gin1, hir1, sas2, c2_05220c_a, fhl1, c2_07690w_a, ash1, rna1, rad50, ycg1, zcf32, ssn6, c4_01660w_a, c4_01780c_a, rfx2, c4_02400c_a, ume7, c4_02570c_a, dot4, srb8, mph1, ybl053, zcf20, orc4, leu3, cst20 med14, c5_04640c_a, ctf5, c6_02430w_a, c6_04060w_a, rfg1, gzf3, cr_02930w_a, yku80, cr_04720c_a, cr_06470w_a, cr_06960w_a, ura2, pif1, ace2, cr_07600w_a, asf1, cr_09490w_a*

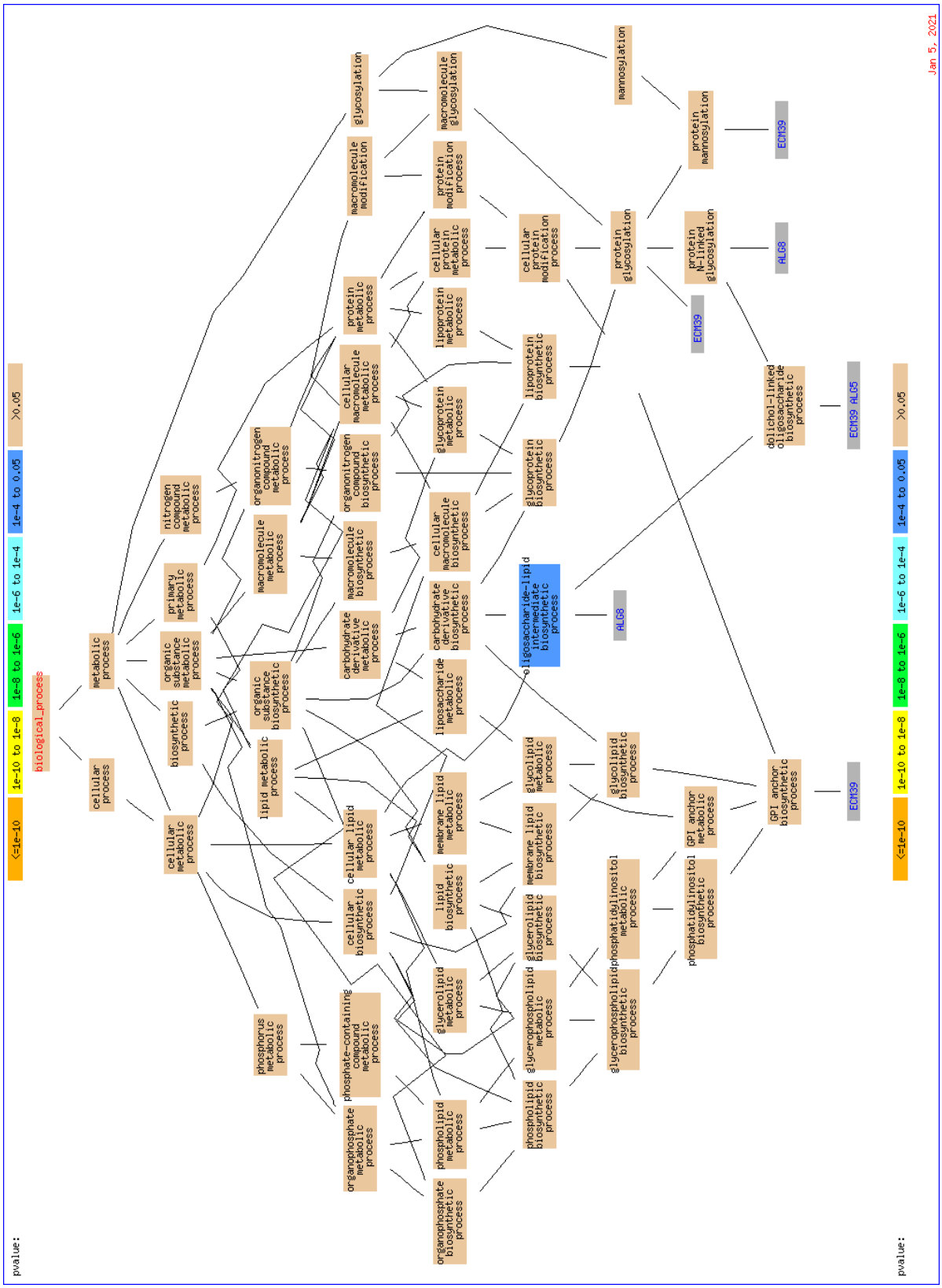
Appendix 2. Graphic displaying the relationships among the GO terms related to molecular function associated to genes with unique missense SNPs in *C. albicans* obtained from vaginal exudates. The analysis was done on January 5th, 2021.



Appendix 2.1. Information regarding genes with unique missense SNPs in *C. albicans* obtained from vaginal exudates, annotated to GO terms from the molecular function ontology.

GO term	Cluster frequency	Background frequency	P-value	Genes annotated to the term
Lysophosphatidic acid acyltransferase activity	3 out of 239 genes, 1.3%	269 out of 6473 background genes, 0.1%	0.04	*c1_04110w_a, c3_00400c_a, lpt1*
Transferase activity, transferring acyl groups other than amino-acyl groups	13 out of 239 genes, 5.4%	117 out of 6473 background genes, 1.8%	0.06	*c1_04110w_a, eco1, pot1-2, c2_06040c_a, lpt1, c3_00400c_a, c3_03760w_a, c4_00810c_a, c4_07210w_a, c6_04410c_a, epl1, cr_03300c_a, cr_08640c_a*
Lysophospholipid acyltransferase activity	3 out of 239 genes, 1.3%	5 out of 6473 background genes, 0.1%	0.09	*c1_04110w_a, c3_00400c_a, lpt1*

Appendix 3. Graphic displaying the relationships among the GO terms related to biological process associated to genes with unique missense SNPs in *C. albicans* obtained from the oral cavity. The analysis was done on January 5th, 2021.



Appendix 3.1. Information regarding genes with unique missense SNPs in *C. albicans* obtained from the oral cavity, annotated to GO terms from the biological process ontology.

GO term	Cluster frequency	Background frequency	P-value	Genes annotated to the term
Oligosaccharide-lipid intermediate biosynthetic process	3 out of 41 genes, 7.3%	18 out of 6473 background genes, 0.3%	0.04	*alg5, alg8, ecm39*