TESI DOCTORAL UPF 2017

Population genomics of the emerging yeast pathogen *Candida glabrata*

Laia Carreté Muñoz

Director

Dr. Toni Gabaldón

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

Als meus pares,
per estar sempre al meu costat.

I en acabat, que cadascú es vesteixi
com bonament li plagui, i via fora!,
que tot està per fer i tot és possible.

Miquel Martí i Pol.

Acknowledgments

He volgut ser biòloga des de que tinc ús de raó. La ciència de la vida, dels éssers vius, del seu orígen i evolució. Entendre, compendre, saber com funciona la vida. Treballar en un laboratori era la resposta que donava quan de petita em preguntaven què volia ser de gran. Vaig estudiar biologia, un màster i aquí estic ara, escribint els agraïments de la meva Tesi. Unbelievable.

Recordo el primer dia que vaig comenar el doctorat, recordo que estava nerviosa i emocionada. Sabia que era un pas important per mi, un inici. Veia que les coses es començaven a posar al seu lloc. Si ara em poso a pensar en els primers dies en el grup, sento que he crescut, no només com a científica, sinó també com a persona, i amb només 4 anys!

Tot això no hauria sigut possible sense l'ajuda del **Toni**, al meu supervisor; qui, sense saber-ho, em va ajudar a construir un camí a seguir. Gràcies, infinites gràcies per tot. Per donar-me la oportunitat de fer el doctorat; per creure en mi; per tenir infinita paciència; per pensar que sí, que sóc capaç de fer i aprendre mil coses; per convertir-me en científica. Gràcies. Et mereixes tots els premis i més, i una plaça fixa al CRG!

Marina i **Cinta**, crec que les dues sou uns pilars en el meu camí. **Marina**, m'has ensenyat a programar i a ser bioinformàtica, has tingut infinita paciència i t'ho agraixo, no saps prou. **Cinta**, m'has ensenyat el món de la genètica de poblacions i m'has fet veure que aquest és el

meu lloc dins de la recerca. Merci pels bons moments dins i fora de la feina. Només dir que ja em sentiré molt afortunada només sent la meitat de bona que vosaltres.

Ernst, Ernestito, don't worry, I will mind the list. See you in Germany, Wien or any place around the world. **Ester**, **Susana** and **Ewa**, the sexy ladies of the wet-lab, thanks for your help, your energy and your smiles. And remember, nobody puts *wetlabers* in a corner! **Irene**, thanks for sharing with me the last crazy days. **Miguel**, King of Murcia, the Freakiest, Khaleesi of ComicSans, the Breaker of Perl, the Father of Fungi, thanks for the funny moments. **Jesse**, thanks for balancing the crazinest in the Colonies. **Veronica** and **Hrant**, it's always nice to have some fresh air and new PhD *candida*tes in the lab. **Rosa**, **Pedro** y **Uciel**, espero que disfrutéis en el lab tanto como yo lo he hecho todos estos años. **Fran**, gracias por tus charlas y ratos, sobre todo cerrando el lab estos últimos días. **Salva**, gracias por todos los buenos momentos y enseñarme qué es (y cómo ser) un work-alcoholic. **Alex**, thanks for having always a smile and a moment to share. **Damian** and **Leszek**, thanks for your programming skills, it was a great support.

Thanks also to all my **friends** in the CRG and the PRBB, we spent great time together during last years. I hope the that our lives will keep crossing in the futur years. **Carla**, **Júlia**, gracias por tantos buenos momentos juntas, por el apoyo mútuo. Espero poder compartir muchos más coffee breaks, ya sea aquí o en la otra punta del mundo. **Imma**, recordo com anava de perduda el primer dia al CRG, gràcies per ajudar-me tant en no perdre'm entre papers i burocràcia. **Romina**, gràcies per ser el nostre àngel de la guarda, no se què fariem sense tu.

And the last, but not least. A tots els meus **amics**, els que han estat, estan i estaran; a les meves amigues del poble, Aida i Ester, la confiana que tinc amb vosaltres no la tinc amb ningú més; als meus amics de la universitat, a les meves companyes de pis, a les més *sepsis* de biologia.

M'és impossible anomenar-vos a tots, vosaltres ja sabeu qui està aquí. No treure suficient nota i anar a fer biologia a Girona ha estat un dels millors errors de la meva vida. Gràcies per estar al meu costat. Sou vosaltres qui ha patit el meu estrés, però també les meves alegries, amb qui he passat dies i nits compartint moments - i espero passar-ne molts més!

I per suposat, a la meva familia, els meus pares i avis, pel suport i la confiança. Vosaltres sabeu que no sempre ha sigut fàcil, no sempre he tingut sort, però tot i així, no heu deixat mai de creure en mi.

I després de tot, de tots aquest anys continuo amb el meu objectiu, amb les ganes de saber més. La cosa no acaba aquí, com diu Queen:

the show must go on.

Laia Carreté Muñoz

Barcelona, September 2017.

Abstract

Infections caused by pathogenic fungi are becoming an increasingly serious threat for human health. Pathogenic fungi such as *Candida glabrata* or *Candida albicans*, belong to phylogenetically distinct clades and have non-pathogenic close relatives, indicating that the ability to infect humans has evolved several times independently. Despite the many recent advances in biomedicine, we are still lacking an understanding of how virulence evolves across organisms and which mechanisms are involved in the emergence of pathogenesis. Elucidating how human pathogens evolve is of central relevance to understand the bases of virulence and spread of infectious agents. In this context, population genomics provides a powerful tool to uncover recent selection pressures that can shed light on how pathogens adapt to humans.

We here evaluated genomic and phenotypic variation across 57 *C. glabrata* strains. Firstly, we focused on 33 globally-distributed isolates. We catalogued extensive copy number variation, which we found to particularly affect genes encoding cell-wall associated proteins, including adhesins. This variation is structured into seven deeply divergent clades, which show recent geographical dispersion and large within-clade genomic and phenotypic differences. We show compelling evidence of recent admixture between differentiated lineages, and of purifying selection on mating genes, which provide first evidence for the existence of an active sexual (or parasexual) cycle in this yeast. Altogether, our results point to a recent global spread of previously genetically isolated populations and suggest

that humans are only a secondary niche for this yeast. Secondly, we analyzed the genomic variability of *C. glabrata* in pairs of serial isolates, each from the same patient. We detected that patients can host clonal and non-clonal isolates. We observed an active standing genetic diversity with recurrent recombination leading to significant differences in terms of oxidative stress resistance biofilm formation. These results suggested that standing genetic variation and within-host recombination between divergent strains may play an important role in disease progression and treatment outcome.

Resum

Infeccions causades per fongs patògens estan esdevenint un greu problema per la salut en humans. La candidiasis, una de les infeccions fúngiques més comunes, està provocada principalment per patògens com *Candida glabrata* o *Candida albicans*. Aquestes dos espècies són filogenèticament distants i tenen altres fongs no patògens al seu voltant, indicant que l'habilitat d'infectar humans ha evolucionat independentment durant els últims anys. Malgrat els avanos en biomedicina, encara estem lluny d'entendre com la virulència ha evolucionat en diferents organismes i quins mecanismes principals han actuat en l'emergència d'aquesta patogenicitat. Entendre com actuen els patògens és de principal importància per entendre les bases de la virulència i expansió d'agents infecciosos. En aquest context, la genòmica de poblacions ens dóna una eina molt valuosa per investigar com la selecció ha actuat en aquests organismes i entendre com els patògens s'han adaptat als humans.

Durant aquest projecte de Tesi, s'ha avaluat genòmicament i fenotpicament 57 soques de *C. glabrata*. Primer, ens hem centrat en l'estudi de la variació genòmica de la població de *C. glabrata* utilitzant 33 soques distribuïdes arreu del món. S'ha catalogat un gran nombre de delecions, duplicacions i aneuploïdies, les quals estan particularment enriquides en proteïnes de membrana o adhesines. Aquesta variació està estructurada en set clades diferents, els quals mostren una recent dispersió geogràfica i una elevada diferència genòmica i fenotípica dins dels clades. L'evidència d'una recent barreja genètica entre diferents clades, i la presència de selecció purificadora en gens rela-

cionats en l'aparellament sexual, proporciona la primera evidència de l'existència d'un cicle sexual actiu en *C. glabrata*. També, s'ha analitzat els canvis genètics de *C. glabrata* en una sèrie de mostres aïllades en diferents dies d'un mateix pacient amb candidiasis. S'ha detectat que els pacients poden tenir soques clonals i soques no clonals. S'ha observat una variació genètica existent i una recombinació recurrent que condueix a diferències significatives en la formació de biofilms. Aquests resultats suggereixen que la variació genètica i la recombinació dins de l'hoste provoca un paper important durant el progrés de la infecció i en el seu futur tractament.

Thesis overview

The overall aim of the present thesis project is to shed light onto the genetic structure and recent evolution of the emerging pathogen *Candida glabrata* by means of sequencing, comparative genomics and population genomics approaches.

This thesis is divided into different chapters, which I briefly introduce here:

Chapter 1 provides an overall introduction to the evolution of pathogenic fungi and the different techniques used in population genomics. **Chapter 2** presents the main objectives of the present thesis.

Chapter 3 is a review of *Candida glabrata* and the emergence of its virulence traits. It provides an overview of recent advances around the question of the evolutionary emergence of virulence traits of *C. glabrata*, using comparisons with *Saccharomyces cerevisiae*, *Candida albicans* and other close relatives.

Chapter 4 presents the analysis of whole genomic variation across 33 representative strains of *C. glabrata* isolated globally. Our comparisons provide a first catalogue of genetic variation, in terms of single nucleotide polymorphisms, gene duplication and loss, as well as gross genomic re-arrangements. This genetic variability is structured into at least seven different clades, which are geographically widespread and present deep genetic divergence. This structure suggests recent worldwide dispersal of genetically differentiated populations, and a

likely recent association to the human host. In this chapter, we also present the first genomic evidence for the existence of an active sexual (or parasexual) cycle and the ability of mating in *C. glabrata*.

Chapter 5 describes the genomic analysis of a sequential series of clinical samples of *C. glabrata*. This chapter introduces the genomic changes in eleven pairs of strains, each isolated from the same patient. This analyses unearths the existence of co-infection by non-clonal isolates. Also describes one case of a recurrent recombination between genetically distinct lineages, leading to significant phenotypic differences in terms of adherence or response to stress.

Chapter 6 and **Chapter 7** present a summarizing discussion of the results obtained during the present thesis project and main conclusions respectively.

Finally, the **Appendix** provides a list of studies in which I have participated during my PhD.

Contents

Part I

Introduction

1 Overall Introduction

This chapter provides an overall introduction to the topics and methods that are relevant for the present thesis. A more comprehensive review on the physiology, pathology and evolution of *Candida glabrata* the organism focus of this study- is presented in the Chapter 3. During the last years we have witnessed an explosion of new knowledge about the biology and evolution of microorganisms, thanks to the application of whole genome sequencing approaches (Scazzocchio, 2014). In this context, the availability of complete genomic sequences of human yeast pathogens and their close relatives helps us to study the mechanisms of virulence; to discover which genes are related to virulence; and to assess how the ability to infect humans has emerged during evolution and how it varies across species (Gabaldón et al., 2016). Part of these developments, particularly for *Candida glabrata*, are described in the next chapter. The further use of genomics techniques to obtain, not a single reference genome for a species, but rather explore the genomic diversity among distinct isolates and strains, enables the investigation of the population structure, epidemiology, and recent evolutionary dynamics in yeast pathogens. Such investigations are the focus of Chapter 4 and 5 of this thesis.

1.1 Evolution of Yeasts

Fungi form one of the largest and most diverse eukaryotic kingdoms, encompassing a variety of organisms such as molds, mushrooms, lichens, smuts, rusts and yeasts. Fungi can be found all over the

globe, and appear in both temperate and extreme environments. They show a remarkably physiological diversity and provide essential contributions to our biosphere, and to human medicine, research and industry, through their functions of decomposing and synthesizing diverse biopolymers, biomolecules and other compounds (Stajich et al., 2009; Petersen, 2013). The origin of the fungal kingdom has been assessed based on the use of molecular clocks, with estimates ranging from 760 million years ago to 1.06 billion years ago (Watkinson et al., 2016). During the last years, researchers have investigated the evolution and diversity of fungi, first using morphological approaches and now, using molecular approaches. The taxonomy of the kingdom Fungi is constantly being changed, especially due to recent research based on DNA sequencing. Currently, eight main phyla are recognized: Opisthosporidia, Chytridiomycota, Neocallimastigomycota, Blastocladiomycota, Zoopagomycota, Mucoromycota, Ascomycota and Basidiomycota (Figure 1.1), with Ascomycota and Basidiomycota representing the subkingdom Dikarya (Hibbett et al., 2007; Richards et al., 2017).

Opisthosporidia is a phylum that includes Aphelida, Microsporidia and Cryptomycota (Karpov et al., 2014). It is an early-diverging group showing primarily a phagotrophic lifestyle. The phylum Chytridiomycota has a worldwide distribution. These fungi and their close relatives, Neocallimastigomycota and Blastocladiomycota, are the only aquatic fungi with active motility, producing uniflagellate zoospores. Chytridiomycota are generally aerobic fungi that can operate as saprotrophs and pathogens in freshwater, marine and soil habitats. Neocallimastigomycota are anaerobic fungi that inhabit the digestive tract of herbivores. As mentioned above, these fungi produce uniflagellate zoospores but multiflagellate spores are also described in some species of the phylum (Hibbett et al., 2007). Blastocladiomycota can live in freshwater, mud and soil, where

they can operate as saprotrophs, decomposing animal trash and plants, or parasitizing arthropods. In contrast to their close relatives, Blastocladiomycota fungi can undergo sporic meiosis (James et al., 2006).

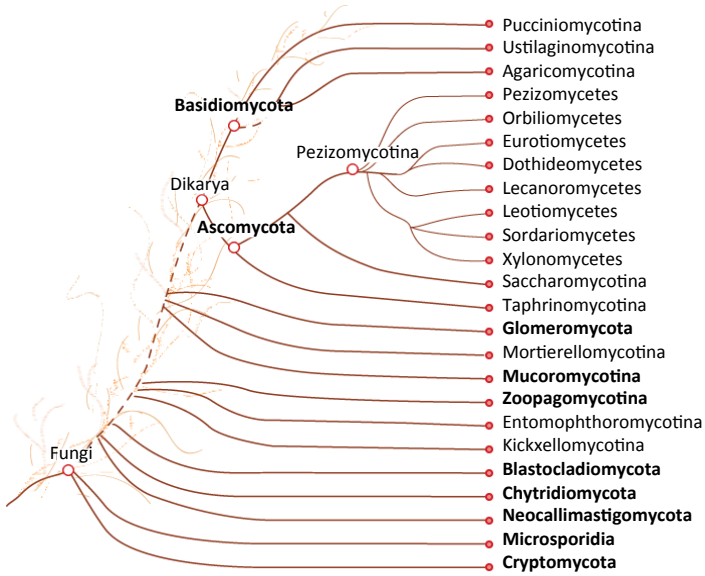

Figure 1.1: Phylogenetic tree adapted from JGI MycoCosm webpage (Grigoriev et al., 2014). The tree illustrates the relationship between major fungal groups. In bold, the main fungal groups explained in text.

Blastocladiomycota, Chytridiomycota and Neocallimastigomycota are fungi with flagellate cells. The lost of flagella is thought to have occurred once in the fungal lineage. The non-flagellated group is composed by Basidiomycota, Ascomycota, Mucoromycota and Zoopagomycota (Hibbett et al., 2007; Spatafora et al., 2016; Richards et al., 2017). Zoopagomycota comprises species that are mainly parasites and pathogens of small animals, for example amoebas, insects, other fungi, etc (Spatafora et al., 2016). Mucoromycota is sister to Dikarya. Mainly consists of plant-associated fungi (Spatafora et al., 2016) and also contains Glomeromycota, although the taxonomic

assignment of this group is yet to be fully clarified. Glomeromycota is one of the main groups of species that establish mutualistic symbioses with plant root cells (Redecker and Raab, 2006; Smith and Read, 2008).

Basidiomycota is one of the two phyla that form the subkingdom Dikarya. Basidiomycota are filamentous fungi that generally form hyphae, with the exception of species that primarily grow in the yeast form such as pathogens of the genus *Cryptococcus*. Ascomycota is the other phylum that constitute the subkingdom Dikarya. Ascomycota is the largest phylum of Fungi that includes morels, mushrooms and unicellular yeasts (e.g. *Saccharomyces*, *Pichia*, *Kluyveromyces* and *Candida*) and many filamentous fungi that live as saprotrophs, parasites and mutualistic symbionts (Watkinson et al., 2016). They generally have sexual structures, however, some species are described as asexual with no known sexual cycle.

Ascomycota phylum is particularly important to humans with many species that can be sources for important medical compounds, such as antibiotics, or are used in food fermentation process, such as in the elaboration of alcoholic beverages, cheese and bread. But also, this clade is an important source of pathogenic yeasts for humans and plants. Human pathogenic yeasts are generally opportunistic pathogens, which can cause infection in susceptible persons, generally those with a compromised immune system.

1.2 Pathogenic yeasts

Many yeast species are part of the human microbiome or are in constant contact with us as part of our environment. Furthermore, several yeast species are used to ferment food or beverages, such as *Saccharomyces cerevisiae*, which is commonly used in making bread or in beer or wine fermentation. Among these many species that are normal components of our microbiota or environment, some

have the particularity of being able to cause disease under certain circumstances, generally those associated to a weakened host immune system. These potentially harmful species are deemed opportunistic yeast pathogens and constitute an important problem of growing medical concern. Indeed, since 1980s, opportunistic yeast pathogens are becoming a major source of life-threatening nosocomial infection, and both their incidence and the number of species involved are growing. This situation is explained in part by recent medical progress. Factors associated to medical progress that favor fungal infections include the extensive use of antibiotics, the use of catheters, the use immunosuppressive chemotherapy, the increased survival of risk groups such as immunocompromised patients, transplant patients, the elderly, or neonates (Pfaller and Diekema, 2007).

Despite recent advances in treating fungal infections, their associated mortally rates remain high at 30-40%, and the treatment is complicated by the existence of resistance to antifungals in some species, and the emergence of novel pathogenic species. The most prevalent infections in humans caused by pathogenic yeasts are those caused by dermatophytes and different species of *Candidas*: mostly *Candida albicans*, *Candida glabrata* and *Candida parapsilosis*, generally in this order attending to their relative incidence (Pfaller and Diekema, 2007; Pfaller et al., 2010; Diekema et al., 2012).

1.2.1 Saccharomycotina subphylum and the distribution of *Candida* species

The *Candida* genus belongs to the Saccharomycotina subphylum, within the ascomycetous fungi (Figure 1.2). Today, more than 1,500 Saccharomycotina yeast species have been described and classified (Kurtzman et al., 2011), and for about 100 of those we have a fully-sequenced genome (Dujon and Louis, 2017), being *Saccharomyces cerevisiae* the first sequenced yeast and the first eukariotic genome available (Goffeau et al., 1996). Plans are underway to sequence over

1,000 species sampled across all described genera with several projects involved in it such as 1KFG project (http://1000.fungalgenomes.org), the 1002 Yeast Genome project (http://1002genomes.u-strasbg.fr), the iGenolevures Consortium (http://gryc.inra.fr) and the Y1000+ Project (http://y1000plus.org) (Hittinger et al., 2015).

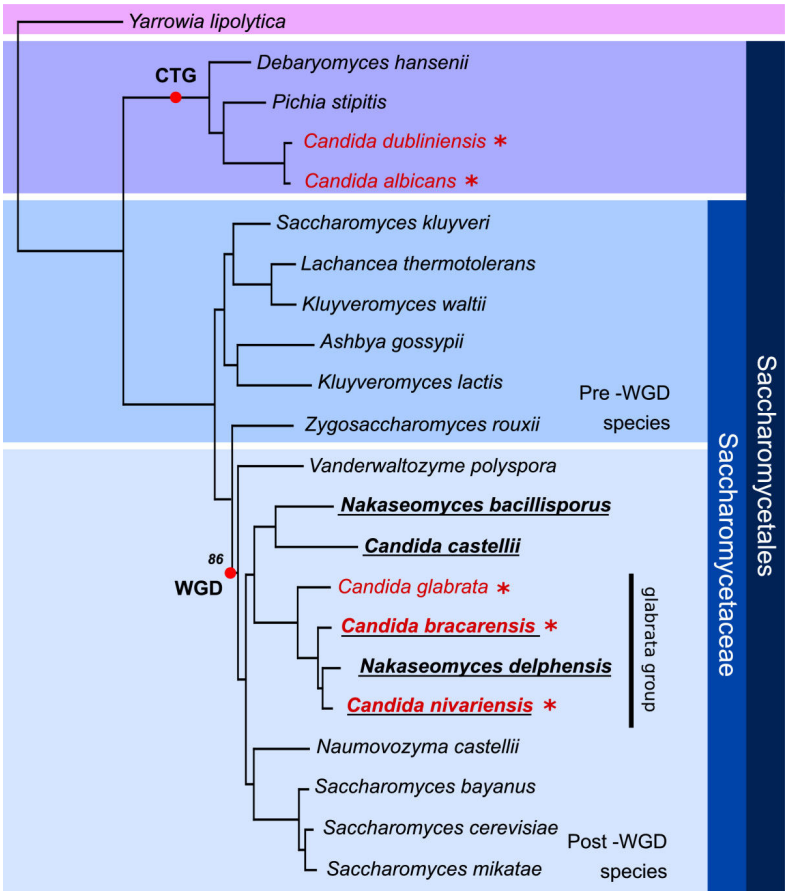

Figure 1.2: Classification of the Saccharomycotina subphylum. Maximum likelihood species tree of 22 Saccharomycotina species from Gabaldón et al. 2013.

The first eight complete genomes sequenced from Saccharomycotina were *Candida albicans* (Jones et al., 2004), *Candida glabrata*, *Kluyveromyces lactis*, *Debaryomyces hansenii*, *Yarrowia lipolytica* (Dujon et al., 2004), *Eremothecium gossypii* (Dietrich et al., 2004) and *Lachancea*

waltii (Kellis et al., 2004). Initial comparisons showed that, *C. glabrata*, *K. lactis*, *E. gossypii*, and *L. waltii*, shared many genomic features with *S. cerevisiae*; grouping them together within the Saccharomycetaceae family. Alternatively, *C. albicans*, *D. hansenii* and *Y. lipolytica*; showed very distinct characteristics, including the usage of an alternative genetic code in *C. albicans* and *D. hansenii* (Sugita and Nakase, 1999; Fitzpatrick et al., 2006).

	Group	Main species	Characteristics	Genome characteristics
Saccharomycotina	early lineages	*Yarrowia lipolytica* *	Lipophilic, the most extensively studied "non conventional yeast" for bioteochnologies	Genome Size: 12 - 24 Mb (> 45% coding) 61000 – 6800 CDS (15-35% split by introns)
	Methylotrophs	*Dekkera bruxellensis*	Food industry	Genome Size: 9 – 13 Mb (> 70% coding) 5000 – 6000 CDS (15% split by introns) Duplicated MAT cassettes
	"CTG" clade	*Candida albicans* *	Human pathogen, asexual, heterozygous diploid	Genome Size: 12 – 14 Mb (> 70% coding) 6100 – 6400 CDS (6-7% split by introns)
		Debaryomyces hansenii *	Osmotolerant, cryotolerant, food industry	Alternative genetic code (CUG > Ser)
	Saccharomycetaceae	*Saccharomyces cerevisiae* *	Most studied fungi, primary reference	Genome Size: 9 – 14 Mb (> 70% coding)
		Candida glabrata *	Human pathogen, asexual, haploid	4700 – 6000 CDS (3-5% split by introns)
		Lachancea waltii *	Unexplored lineage, non-duplicated genome	Triplicated MAT cassettes
		Kluyveromyces lactis *	Lactose utilizing yeast, non-duplicated genome	Point centromeres
		Eremothecium gossypii *	Plant pathogen, non-duplicated genome	No complex I genes in mtDNA

Table 1.1: Saccharomycotina groups and their principal characteristics. First column indicates name of groups. Second column with a list of the main species, $*$ indicates the first eight fully sequenced yeast genomes published in year 2004 (original references in text). Third and fourth column indicates main characteristics and genome characteristics respectively. Table adapted from Dujon and Louis 2017

More recent studies that use dozens of sequenced genomes recognize four major groups within the Saccharomycotina subphylum (Table 1.1). The first one, and the most extensively studied family, is the Saccharomycetaceae; second, the CTG clade, which comprises yeasts

that use an alternative genetic code; third, the methylotroph clade; and fourth, a heterogeneous group which comprises different species belonging to distant lineages regarded as early diverging within the Saccharomycotina subphylum.

The Saccharomycetaceae family is the most studied family, having now at least one complete genome sequenced for each main lineage. Members of this family share triplicated MAT cassettes, the absence of complex I subunits of the respiratory chain in their mitochondrial DNA (mtDNA) and the presence of point centromeres. The CTG clade has a defining characteristic that clearly distinguishes it from the other groups: the use of an alternative genetic code, in which the CUG codon codes for serine instead of leucine (Santos et al., 2011). Other characteristics of this clade are the presence mitochondrial genes encoding subunits of complex I of the respiratory chain and the presence of a single mating-type locus. The Methylotroph clade shares almost all characteristics with the CTG clade, with the exception of the usage of the alternative genetic code. In addition genomes of species within this clade are moderately smaller (Table 1.1). The remaining group comprises several early divergent lineages within the Saccharomycotina. The genomes of the species in this clade are significantly larger and display lower compactness in terms of protein-coding genes, as compared to those species of the other clades (Table 1.1) (Dujon and Louis, 2017).

1.2.2 Distant cousins: *Candida albicans* and *Candida glabrata*

The present thesis has focused on analyzing genomic features across sequenced *Candida glabrata* isolates. During the following lines and the following chapter, we compare the main characteristics of *Candida glabrata* to *Candida albicans*, as the latter is considered the main model organism for Candida pathogenesis. As mentioned above *C. albicans*, and *C. glabrata* are the first and second most prevalent species in candidemias (i.e. invasive infections caused by *Candida* species),

respectively, and are phylogenetically distantly related, belonging to two different clades within Saccharomycotina. Several previous studies have established differences and similarities between these two yeast pathogens. Although *C. albicans* is still the most common pathogenic *Candida* species, the last years have seen an increase in the incidence of non-albicans *Candida* species in hospital-acquired infections, with *C. glabrata* becoming increasingly relevant (Gabaldón et al., 2016). This could be partially attributed to the capacity of *C. glabrata* to acquire resistance to commonly used antifungals (Fidel et al., 1999; Pfaller and Diekema, 2007; Pfaller et al., 2010; Diekema et al., 2012; Vale-Silva and Sanglard, 2015; Perlin et al., 2015).

Candida glabrata presents important phenotypic differences with *C. albicans*. *C. glabrata* is strictly haploid, which is in contrast to *C. albicans* that is an obligate diploid. Until now, no sexual cycle has been documented in *C. glabrata* although genes involved in sexual reproduction are known to be conserved (Wong et al., 2003; Srikantha et al., 2003; Fabre et al., 2005). Only recently, it has described that *C. glabrata* can recombine and perform mating type switching (Dodgson et al., 2005; Carreté et al., 2017). Another important difference which relates to virulence is the ability to form hyphae, which is present in *C. albicans* but absent from *C. glabrata*. Certainly, switching to hyphal growth is a recognized virulence mechanisms in *C. albicans*, which allows this yeast to penetrate host tissues and escape from macrophage engulfment (Mayer et al., 2013). In contrast to the aggressive strategies used by *C. albicans*, *C. glabrata* evades the immune system, and can survive within macrophages, where it can even proliferate (Kaur et al., 2007; Roetzer et al., 2010; Brunke and Hube, 2013).

Infections caused by *Candida* yeasts are typically treated with azole and echinocandin drugs. Continuous exposure to such agents often leads to drug resistance. Previous studies have identified different

mechanisms for *Candida albicans* and *Candida glabrata* to acquire resistance to drugs. Those mechanisms principally affect the activity pathways of target proteins to acquire resistance to drugs, positioning *ERG11*, *CDR1*, *CDR2*, *FKS1*, *FKS2*, *PDR* gene family and *MDR1*, as the principal ones (Asai et al., 1999; Coste et al., 2005; Oliver et al., 2007; Dunkel et al., 2008; Katiyar et al., 2012; Vale-Silva and Sanglard, 2015; Perlin et al., 2015; Demuyser et al., 2017). Gene *ERG11* has a role in ergosterol biosynthesis and is a target of azole antifungals. *CDR1* and *CDR2* are genes involved in multidrug transporter of ATP-binding cassette (ABC) superfamily and are over-expressed in azole-resistant isolates. *FKS1* and *FKS2* are an essential glucan synthase subunits, an allelic variation determine resistance or sensitivity to echinocandin drugs. *PDR* family encodes proteins that regulate drug efflux pumps. *MDR1* is also related to multi-drug efflux pump and its over-expression confers fluconazole resistance.

It is thought that the emergence of resistance to antifungal agents is favored by a high plasticity at the genomic level. Such resistance related genome alterations can consists of different single mutations in genes encoding these proteins or caused by copy number variation of these genes, including genome rearrangements and aneuploidies (Marichal et al., 1999; Selmecki et al., 2006, 2008, 2009, 2010; Ford et al., 2015). Beyond the resistance to antifungal drugs explained above, *Candida glabrata* possess other abilities that place it in the spotlight of pathogenic fungi. *C. glabrata* genome encodes a particular family of epithelial adhesin proteins called EPA, which were considered a key factor in the ability to infect humans (Cormack and Falkow, 1999; Roetzer et al., 2011; Gabaldón et al., 2013; Vale-Silva et al., 2017).

1.2.3 Sexual, asexual and parasexual cycle

Mating-type is a molecular mechanism that regulates compatibility in haploid cells. The genomic recombination mechanisms that enables changing from one mating-type to another one is called mating-

type switching. Mating-type switching its done by interchanging the content of mating type (*MAT*) locus through genomic recombination (Butler et al., 2004; Haber, 2012; Hanson and Wolfe, 2017).

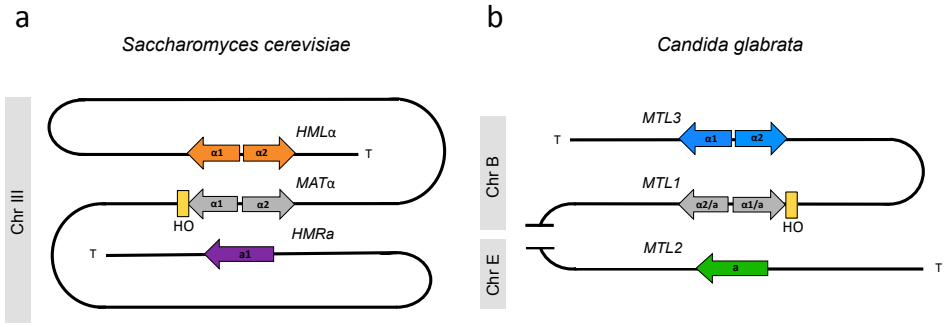

Figure 1.3: Mating type loci of *S. cerevisiae* and *C. glabrata*. a) Organization of mating type loci in chromosome III in *S. cerevisiae*: *MAT* locus marked in grey, *HMR* in purple and *HML* in orange, respectively. Cutting zone from *HO* enzyme marked in yellow. b) Structure of mating type loci in *C. glabrata*. *MTL1* in grey, *MTL2* and *MTL3* marked in green and blue respectively. Yellow bar indicates cut zone from HO enzyme.

Saccharomycetaceae present three mating type loci, in contrast to *C. albicans* from CTG clade that only have one genomic locus encoding the mating genes (Hanson et al., 2014). *C. albicans* presents a parasexual cycle (Hull and Johnson, 1999; Forche et al., 2008; Bennett, 2015), in which cells of opposing mating type fuse but do not undergo meiosis. Rather, they return to normal ploidy through differential chromosomal loss. The most studied switching mechanism in yeast is the one from *S. cerevisiae* (Haber, 2012). Mating type loci of *S. cerevisiae* are all located in chromosome III, and are named *MAT* locus, *HMRa* and *HMLalpha*. The gene conversion is initiated by a double-strand break made by the *HO* enzyme. During the process, *MAT* locus is the only locus that can be cut and only one of the two cassettes (*HMLalpha* or *HMRa*) is expressed. The other cassette is shaded from the *HO* enzyme by a silencing process mediated by *Sir* proteins. Then, *HMLalpha* or *HMRa* cassette are used as a template to repair the *MAT*

locus by homologous recombination (Figure 1.3a).

Similar to *S. cerevisiae*, *C. glabrata* genome encodes both mating types and has homologous genes for all the genes involved in mating in *S. cerevisiae* (Muller et al., 2008). In *C. glabrata*, the three loci are called *MTL1*, *MTL2* and *MTL3* which correspond to *MAT* locus, *HMR*, and *HML*, respectively. *MTL1* contains *MAT* locus, *MTL2* contains **a** information, and *MTL3* contains alpha information (Figure 1.3b) (Yáñez-Carrillo et al., 2014). In contrast to *S. cerevisiae*, the three loci are in different chromosomes with *MTL1* and *MTL3* being located in chromosome B and *MTL2* in chromosome E. Mating type switching has been suggested by PCR experiments (Butler et al., 2004), and recently it has described that *C. glabrata* can recombine and perform switching (see Chapter 4) (Dodgson et al., 2005; Carreté et al., 2017). But despite great efforts, no spontaneous switching has been observed in laboratories, as it is the case in *S. cerevisiae*.

1.3 Introduction to population genomics

Population genetics has been a discipline of interest for many scientist. At the beginning of the 20th century, population genetics was mostly a theoretical discipline, in part due to the scarcity of empirical data. With the advent of DNA sequencing, the scale of empirical data has been growing exponentially during the last decades. However, for the most part, this growth has been limited to certain marker genomic loci. The evolution of techniques such as the polymerase chain reaction (PCR) and Sanger sequencing enabled the production of genetic data that has been extremely useful to confirm theoretical data and increase our knowledge on the population structure and dynamics of species of interest. However, the use of selected loci still provided only a partial view of the entire genetic complement and a limited resolution of the processes under study.

In the last two decades, the advent of high-throughput sequencing

techniques, coupled with parallel progress in computer architecture and bioinformatics approaches has paved the way to interrogate the full genetic complement of a growing number of individuals of the species of interest. This has enabled the evolution from population genetics to so-called population genomics. Population genomics can be defined as the large-scale comparison of the whole genetic information of different populations. Linked to population genetics and their limited study of a small amount of loci, population genomics can improve our understanding of the evolution, molecular ecology and epidemiology of pathogens (Luikart et al., 2003; Pool et al., 2010).

Population genomics provides a solid background for understanding the distribution of genetic variability among different populations, and a support to understanding the contributions of mutations, natural selection and genetic drift in the evolution of genes and genomes. In a population, there are many causes that can develop genetic variability, for example: recombination, polyploidization, migration between populations coupled to genetic admixture, population expansions, and genetic mutations (Charlesworth, 2010). Genetic variability comprises genetic variants in coding and non-coding regions, and can range from single nucleotide polymorphisms to large structural changes at the chromosome level. Those alterations can produce no negative or positive effect on the phenotype (i.e. neutral changes) or, in contrast, produce changes at the phenotypic level, which in turn will affect fitness and be positively or negatively selected (Conrad and Hurles, 2007; Gasperskaja and Kučinskas, 2017).

1.3.1 Genomic variation and evolutionary processes

Genomic variants can consist of single-nucleotide polymorphisms (or SNPs) or affect more than a single nucleotide, affecting an entire fragment of a sequence and generate a deletion or duplication (defined as a copy number variation or CNVs) (Figure 1.4). Further types of variants include changes in the relative order or orientation in

the sequence such as inversions, translocations and other forms of re-arrangements, which are generally generated though recombination (Figure 1.4a).

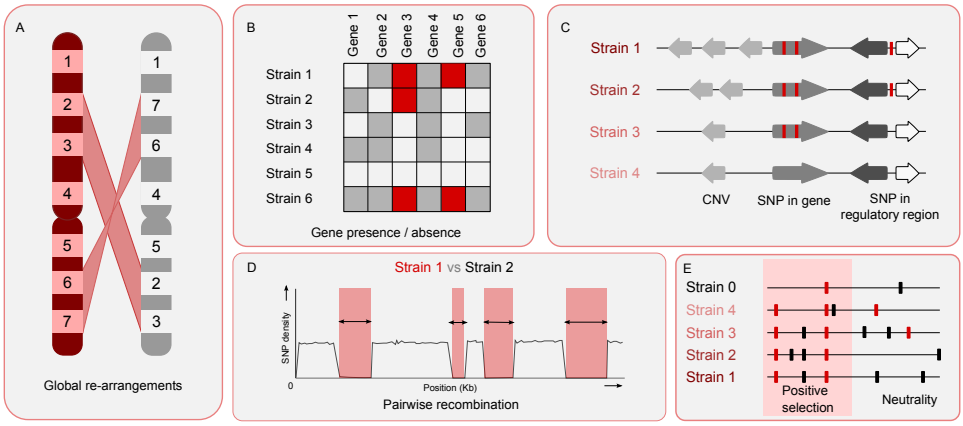

Figure 1.4: Schematic overview of comparative genomics methods adapted from Gabaldón et al. 2016. a) Search for genomic re-arrangements. b) Presence and absence of genes. Heatmap represent duplications (in red) and deleted genes (in white). c) Detection of differences between genomes. Each arrow represent a gene. Red horizontal lines represent SNPs. d) Pairwise recombination between two genomes. Y-axes represent SNP density and X-axes represent genome position. Red zones with arrows indicates fragment recombined. e) Detection of positive selection. Black horizontal lines indicates synonymous SNPs and red horizontal lines represent non-synonymous SNPs.

Recombination is the process by which genomic segments from different parts of the genome are combined, creating a new sequence that differs from the pre-existing ones. Recombination can occur between homologous regions in a genome, e.g. between homologous pairs of chromosomes in diploid organisms, or between non-homologous regions of the same or distinct chromosomes. Depending on the type of recombination the new variants would result in translocations, inversions, deletions, duplications or gene conversion (i.e. the substitution of a given genomic fragment by an equivalent one copied over from a homologous region). Of note recombination after crossing of individuals from different populations leads to genetic flux and to assortment of variants from the two populations in the same chromo-

some. Recombination estimates are key to understand the relationships between recombination and population genomics parameters. For example, variations of the recombination rate across the genome can reveal selection processes. Recombination can also be detected using pairwise comparisons between genomes by spotting the presence of large regions without SNPs between the compared genomes (Figure 1.4d). For an ancestral shared variation we could expect to find SNPs similarly distributed across the whole genome, for a recombination event, in contrast, we could expect large blocks of regions without SNPs, corresponding to the recombined fragments.

The non-random association of alleles at different loci in a population is called linkage disequilibrium (LD) (Slatkin, 2008). LD is influenced by different and independent factors including the different types of selection, rate of recombination, rate of mutation, population structure and genetic drifts (Conrad and Hurles, 2007). Recombinations play an important role in shaping patterns of LD in a population. When recombination occurs in one part of the genome, it tends to reduce the dependence between those sequences and thus reduce LD. As a result, the pattern of LD in a genome is a footprint of the genetic processes affecting a population and its genetic structure (Li and Stephens, 2003; Slatkin, 2008).

Genetic polymorphisms of any type can affect the function or expression levels of different genes and, as a consequence, have an impact on the phenotypic traits of the individual. When such mutations reduce the fitness of the carrying individuals, natural selection may reduce the frequency of such variants in the population, or eventually, remove them completely. This process is known as purifying selection. In contrast, if the final trait caused by a particular variation is favorable for the individual as compared to others in the population, natural selection will tend to increase the frequency of those variants, eventually leading to fixation. This

process is known as positive selection (Figure 1.4e). Positive selection can be represented by different models, for example: the classic hard-sweep model, when a mutation appears in one individual and increases the frequency rapidly in the population; standing variation, where selection acts on a previously existing mutation by increasing its frequency; and polygenetic adaptation, where simultaneously occurring mutations at different parts of the genome increase in frequency in the population.

It is important to note that processes of natural selection depend on certain parameters such as population size, structure and dynamics. Demographic processes such as population growth or a decrease of the population size can affect the genetic diversity of a population. The importance of migration comes from the fact that many species are composed by different subpopulations, connected by occasional migration, but largely isolated from each other.

Migration of subpopulations give rise to gene flow, that is the exchange of genetic material between populations. When genetically different populations develop descendants thought migration, the genome in the resulting population contains genetic material from both of the ancestral populations. This process of genetic mixing is called admixture and results in an admixed population. On the other hand, mating between different species can drive also to the formation of hybrid organisms. When the genetic material is transferred across species this results in reticulated patterns of evolution that include vertical and non-vertical modes of evolution.

1.4 Genomic era and techniques

1.4.1 Large-scale sequencing

Since 1990s, several nucleotide sequencing methods were developed that are based on the chain-termination method developed by Fred-

eric Sanger and colleagues (Sanger et al., 1977; Swerdlow et al., 1990; Hunkapiller et al., 1991). Since mid-2000s, great advances in large-scale sequencing techniques using different approaches developed quickly. These techniques were coined with the generic name Next generation sequencing (NGS) techniques (Shendure and Ji, 2008; Shendure and Aiden, 2012; van Dijk et al., 2014). NGS techniques share three majors characteristics: they (i) are based on the preparation of sequencing libraries in a cell free system, (ii) are produced in parallel and (iii) the output is directly detected without the use of electrophoresis. The emergence of NGS techniques, immediately revolutionized genomics research by bringing about the possibility of sequencing entire genomes at reasonable costs. Now, whole-genome sequencing (WGS) has been popularized and pervades many areas of biological research, due to the declining cost of sequencing and the extensive availability of sequencing platforms.

1.4.2 Whole-genome sequencing analysis

Thanks to the availability of thousands of genomes we can easily compare different species and obtain information of the genetic variation existing within a species. Genotyping is a process to determine genetic differences of a sample by analyzing the DNA sequence and comparing it to another sample or a reference sequence. The measurement of genetic variations in terms of single nucleotide polymorphism (SNPs), deletions and duplications, or large sequences (e.g. aneuploidies) can provide a view of how genetic variation evolves and how a given population was structured over time. Different approaches, such as phylogenetic trees based on WGS, statistics to evaluate the distribution of SNP (e.g. Principal component analysis), model-based clustering methods (e.g. Structure or Admixture software), can help us to discover the genetic structure and recent evolution of a population.

The availability of multiple genomes from within a given species,

brings the need of efficient tools that enable their comparison. Softwares such as Mauve (Darling et al., 2004, 2010), Mugsy (Angiuoli and Salzberg, 2011), MAFFT (Katoh et al., 2002), among others, are helping to analyze huge amount of genomes as we used in Chapter 4. Principal component analysis (PCA) or Multiple correspondence analysis (MCA) are used to analyze genetic diversity, in this case based on SNPs data, using matrix approach in a non-categorical and categorical data, respectively (Price et al., 2006; Patterson et al., 2006). Model-based clustering methods assume a model in which there are an unknown number (K) of ancestral populations, each characterized by a set of allele frequencies, that contribute to the genetic background of the analyzed individuals (Pritchard et al., 2000; Falush et al., 2003, 2007; Hubisz et al., 2009). Applications of this method include, among others, the assessment of population structure, identification of genetically distinct populations, identification of migrant and admixed individuals, and clustering of individuals within populations.

1.5 Phylogenomics

Understanding the relationships between organisms using phylogenetics is a basic step of almost any evolutionary study. The idea of phylogenetics stems from the theory of evolution presented in the famous book The Origin of Species by Charles Darwin (Darwin, 1859). Phylogenetics relies on using different methods to understand the evolution of past and recent species and provide an overview of the evolutionary history of life on Earth. Until 1970s, phylogenetic reconstruction was based on morphological traits and ultra-structural characters. However, this approach cannot be applied to microorganisms, as the number of differences at the morphological level are very limited. The introduction of molecular data and its use to reconstruct evolutionary relationships represented a revolution (Zuck-

erkandl and Pauling, 1965). Indeed, the access to DNA sequences generated an increased number of homologous sequences that could be compared, and the number of evolutionary reconstructions for different groups proliferated. The comparison of homologous sequences shared among species, serves to reconstruct phylogenetic trees and infer past evolutionary processes. Two homologous sequences of DNA can have shared ancestry because of a speciation event (orthologs) or because of a duplication event (paralogs) (Koonin, 2005). There exist three main kinds of reconstruction methods that can be used to infer phylogenetic trees: distance methods, that converts sequence differences into a distance matrix that represents the evolutionary distances between species; Maximum parsimony, that selects the tree with the minimum changes to explain the observed data; and Probabilistic methods (e.g. Maximum Likelihood or Bayesian analysis), which are based on probabilistic functions that model the likelihood that a given phylogenetic tree could have produced the observed sequence data (Delsuc et al., 2005; Bleidorn, 2017).

The huge amount of data generated during the last years opens a window to a new field of research: phylogenomics. Phylogenomics can be regarded as the intersect between the fields of genomics and evolution (Eisen and Fraser, 2003). The term is commonly used to refer to analyses that involve genome data for the reconstruction of evolutionary relationships or processes. Reconstruction methods have been adapted to be used with entire genomes and can be divided into sequence-based methods, and whole genome content methods. Sequence-based methods are based on the comparison of primary sequences, and the phylogenetic tree are based from multiple sequence alignments. In contrast, whole genome-based methods are based on the comparison of whole-genome content data. Those methods based on gene order and gene content do not rely on a multiple-sequence alignment step, but they still depend on homology

and orthology analysis (Delsuc et al., 2005).

There are different research questions hat can be addressed by phylogenomics, including (i) prediction of gene function based on sequence similarity (Eisen and Fraser, 2003), (ii) the identification of events in gene evolution such as gene deletion or duplication, (iii) the identification of horizontal gene transfers, hybridizations, and other forms of reticulated evolution (Whitaker et al., 2009), and (iv) clarification of evolutionary relationships (Delsuc et al., 2005; Philippe et al., 2005; Jeffroy et al., 2006).

2 Objectives

The main aims of the present thesis are to:

- Assess the genetic structure of *Candida glabrata*, seeking for patterns that can inform of a recent or ancient association with the human host.
- Assess evolutionary dynamics and plasticity of *C. glabrata* genomes and relate it to their phenotipic effects.
- Search for genomic footprints of active mating, recombination, and mating type switching systems in *C. glabrata*.
- Reconstruct the recent evolution of *C. glabrata* populations with a focus on genetic diversification and admixture.
- Assess levels of genetic diversity and evolutionary processes in serial *C. glabrata* isolates obtained during the course of an infection.
- Establish the relevance of different mutation mechanisms (single-nucleotide polymorphisms, aneuploidies, re-arrangements, copy-number variation) in *C. glabrata* and their possible role in the adaptation to the human host.

Part II

Results

3 The bird of a deadly yeast: tracing the evolutionary emergence of virulence traits in *Candida glabrata*

Gabaldón T, Carreté L. The birth of a deadly yeast: tracing the evolutionary emergence of virulence traits in Candida glabrata. FEMS Yeast Res. 2016 Mar;16(2):fov110. DOI: 10.1093/femsyr/fov110

4 Patterns of genomic variation in the opportunistic pathogen *Candida glabrata* suggest the existence of mating and a secondary association to the human host

Carreté L, Ksiezopolska E, Pegueroles C, Gómez-Molero E, Saus E, Iraola-Guzmán S, 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 Jan 8;28(1):15–27.e7. DOI: 10.1016/j.cub.2017.11.027

5 Genomic analysis of serial clinical isolates of *C. glabrata*

Laia Carreté, Ewa Ksiezopolska, Emilia Gómez-Molero, and Toni Gabaldón. (2017) Genomic analysis of serial clinical isolates of C. glabrata suggests that mating and genetic introgression can occur during the course of infection. (In preparation)

5.1 Abstract

Candida glabrata is an opportunistic fungal pathogen of increasing incidence, which currently ranks as the second most common cause of candidiasis. Although the mechanisms underlying virulence and drug resistance in *C. glabrata* are starting to be elucidated, we still lack a good understanding of how this pathogen adapts during the course of an infection. An outstanding question is whether the observed genomic plasticity of *C. glabrata* plays a role during infection. To assess this, we here compare the genomes of serial clinical isolates obtained from the same patients. Our results provide a catalogue of single-nucleotide variations among clonal pairs of isolates, and uncovers an enrichment of non-synonymous changes in genes encoding cell-wall proteins. We show evidence for co-infection with non-clonal isolates, and show, for the first time, that recombination among genetically distant clades can contribute to the genetic variability of infecting isolates. This suggests that genetic admixture between co-infecting or co-colonizing *C. glabrata* strains may play a role in infection and clinical outcome.

Keywords: *Candida glabrata*, recombination, mating, human fungal pathogens, adhesion

5.2 Introduction

Infections caused by fungal pathogens are becoming an increasingly serious medical problem. It is estimated that invasive fungal infections can kill around 1,5 million people every year (Brown et al., 2012b). Many fungal infections are caused by opportunistic pathogens that can also live as normal components of the microbiota of most healthy humans (Cui et al., 2013), but that can cause life-threatening complications when the immune system is weakened. The incidence of opportunistic fungal pathogens has increased during the last years, partly owing to medical progress. Factors contributing to this increase include, among many others: extensive use of antibiotics, increased survival of immunocompromised patients, increased use of invasive clinical procedures (such as the use of catheters, neonatal intensive care or organ transplantation), and the use of immunosuppressive chemotherapy (Pfaller and Diekema, 2007). *Candida* species are the most common source of hospital-acquired invasive infections (Richardson and Lass-Flörl, 2008). Among pathogenic *Candida* species, the most prevalent in human infections are *Candida albicans* and *Candida glabrata*, usually in this order. Antifungal drugs - i.e. azoles, polyenes, echinocandins, among others (Vale-Silva and Sanglard, 2015; White et al., 2002, 1998) - are available to treat infections caused by these species. However, the efficacy of these drugs is limited, owing to late or imprecise identification of the invasive agent and the increased presence of resistance (Pfaller et al., 2003, 2009; Lockhart et al., 2012; Cleveland et al., 2012). Development of antifungal resistance during treatment is a possible cause of treatment failure, but we lack a comprehensive understanding of how this or other adaptive processes develop during the course of an infection, or what is the genetic diversity existing in pathogen populations infecting the same patient.

In this context the study of co-isolates or serially-sampled isolates obtained during the course of an infection can serve to trace variations at the genetic or physiological levels that are of relevance to understand disease and treatment outcomes. Next generation sequencing technologies now allow tackling this from the perspective of the entire genomic sequences of isolated strains. Previous genomic studies with *Candida albicans* serial isolates

identified several alterations than may have contributed to drug resistance in the course of a treatment, including segmental aneuploidies (Selmecki et al., 2006), loss of heterozygosity (LOH) in large parts of the chromosomes (Dunkel et al., 2008), and alterations at the gene level. However, these studies could not differentiate between mutations occurred during the course of infection or selection over pre-existing standing variation. Despite the usually clonal nature of clinical isolates obtained from the same patient, it is known that there exist significant variation within *C. albicans* samples, suggesting that selection to become more resistant to drugs can shape the genetic variation of infecting populations (Ford et al., 2015; Hirakawa et al., 2015).

Candida glabrata possess remarkable differences with *C. albicans* (Gabaldón and Carreté, 2016). These include the fact that *C. glabrata* is a haploid organism and thus LOH does not play a role. In addition *C. glabrata* presents different infection strategies and antifungal properties as compared to *C. albicans*. *C. glabrata* genome encodes several members of a specific family of epithelial adhesin proteins called EPA, which are considered a key factor in the ability to infect humans (Cormack and Falkow, 1999; Roetzer et al., 2011; Gabaldón et al., 2013; Vale-Silva et al., 2017). Another ability that has been confirmed recently, using comparison of whole genome sequences is that genetically diverse lineages can recombine leading to genetic admixture (Dodgson et al., 2005; Carreté et al., 2017). This recent study have showed that *C. glabrata* isolates can be ascribed to at least seven genetically differentiated clades that do not follow a geographical structure (Carreté et al., 2017). Importantly, these studies also revealed that *C. glabrata* is likely able to undergo mating (Carreté et al., 2017). Although such studies show evidence for relatively recent recombination (i.e. unique to one or few isolates within a clade), it is unknown whether this recombination can occur in the course of human infection or commensalism. Several recent studies have compared genome sequences from *C. glabrata* isolates obtained from the same patient (Carreté et al., 2017; Håvelsrud and Gaustad, 2017; Biswas et al., 2017; Vale-Silva et al., 2017). Most of these analyses reveal very little genetic variation, supporting the idea that a single clonal population

colonize different body sites and leads to infection.

Here, we reanalyzed previously available genomic data of ten pairs of *C. glabrata* isolates from the same patient. In addition, we sequenced a trio of serially-isolated strains from the same patient obtained in the course of a week of acute infection. Our aim was to identify common trends in the genomic variation found across isolates from the same patient, with a focus on processes related to disease and drug resistance. Our analyses revealed two unexpected results. On the one hand, two of the eleven isolate pairs from the same patient were shown to be non-clonal but genetically related isolates. On the other hand, we found compelling evidence that strains serially isolated from the same patient show differential signs of genetic introgression from genetically-distant *C. glabrata* clades. This strongly suggest that mating and genomic recombination between co-colonizing strains in the same patient can occur during the course of infection or colonization. Altogether, our results raise new questions regarding the potential contribution of genetic recombination between strains to infection and survival to treatment.

5.3 Results and discussion

5.3.1 Genomic variation in serial clinical isolates

To assess genomic variability present in serial *C. glabrata* clinical isolateswe obtained 20 available datasets from whole genome shotguns corresponding to ten different isolate pairs, each obtained from the same patient over the course of 38 days, in average (Carreté et al., 2017; Håvelsrud and Gaustad, 2017; Biswas et al., 2017; Vale-Silva et al., 2017). In addition, we sequenced three serial isolates (SAT01BAL, SAT02PL, SAT03BC, referred to as SAT strains collectively) obtained from different body sites over the course of a week from the same leukemia patient suffering candidiasis. The three strains were sequenced using Illumina pair-end technology to an average coverage ranging from 119 to 175x. This represents the first trio of serial isolates sequenced to completion. In total 23 strains corresponding to eleven different patients (Table 5.1) were analyzed in a common analytical

framework, to avoid analytical biases. For all strains we used a read-mapping strategy against the reference genome sequence of strain CBS138 (Dujon et al., 2004) and assessed genome variation in terms of single nucleotide polymorphisms (SNPs), copy number variations (CNVs) and genomic re-arrangements (See Materials and Methods).

Strains	Time	Site	Country	Source data
B1012M	0	Oral	Belgium	Enache-Angoulvant *et al.*, (2010)
B1012S	0	Stool	Belgium	Enache-Angoulvant *et al.*, (2010)
EB101M	0	Oral	Belgium	Enache-Angoulvant *et al.*, (2010)
BO101S	0	Stool	Belgium	Enache-Angoulvant *et al.*, (2010)
P35_2	0	Oral	Taiwan	Lin *et al.*, (2007)
P35_3	90	Oral	Taiwan	Lin *et al.*, (2007)
CANGA1A	0	Blood	Norway	Hvelsrud *et al.*, (2017)
CANGA1B	90	Blood	Norway	Hvelsrud *et al.*, (2017)
CANGA2A	0	Blood	Norway	Hvelsrud *et al.*, (2017)
CANGA2B	90	Blood	Norway	Hvelsrud *et al.*, (2017)
CANGA3A	0	Blood	Norway	Hvelsrud *et al.*, (2017)
CANGA3B	90	Blood	Norway	Hvelsrud *et al.*, (2017)
CMRL1	0	Blood	Australia	Biswas *et al.*, (2017)
CMRL2	21	Blood	Australia	Biswas *et al.*, (2017)
CMRL3	0	Blood	Australia	Biswas *et al.*, (2017)
CMRL4	30	Blood	Australia	Biswas *et al.*, (2017)
CMRL5	0	Pelvis	Australia	Biswas *et al.*, (2017)
CMRL6	12	Urine	Australia	Biswas *et al.*, (2017)
SAT01BAL	0	BAL	France	This project
SAT02PL	1	PL	France	This project
SAT03BC	6	Blood	France	This project
DSY562	0	Oral	Switzerland	Vale-Silva *et al.*, (2017)
DSY565	50	Oral	Switzerland	Vale-Silva *et al.*, (2017)

Table 5.1: Information about *C. glabrata* isolates. Columns in the table indicates: Strains; Time: time elapsed between samplings (in days); Site: isolation site; Country; Source. BAL = Bronchiolo-alveolar lavage, PL = Peritoneal fluid

To provide a global comparison of these genomes with previously sequenced isolates from around the globe (Carreté et al., 2017), we ana-

lyzed the obtained SNP patterns using Multiple Correspondence Analysis (MCA). This analysis provided consistent results with previously published clades, and suggested that all these new strains could be ascribed to several of the previously described clades (Carreté et al., 2017): namely, pairs B1012M/B1012S and EB101M/BO101S belong to clade I; SAT strains to clade II; DSY562/DSY565 to clade III; P35_2/P35_3, CMRL1/CMRL2 and CANGA3A/CANGA3B strains belong to clade V; and CMRL3/CMRL4, CMRL5/CMRL6, CANGA1A/CANGA1B and CANGA2A/CANGA2B strains belong to clade IV (Supplementary Figure 5.5). Thus, in this survey all isolates from the same patient belonged to the same genetic clade.

Compared to the reference genome, we observed an overall range from 2.97 SNPs/Kb in CRML6 to 6.53 SNPs/Kb in B1012M. Most of these SNPs were shared between each pair of strains. Overall, we detected small differences between the two genomes of a pair, with the smallest value 0.049 SNPs/Kb between CANGA3A/CANGA3B, and with the highest value 0.776 SNPs/Kb between CMRL5/CMRL6 (Figure 5.1). Of note, the amount

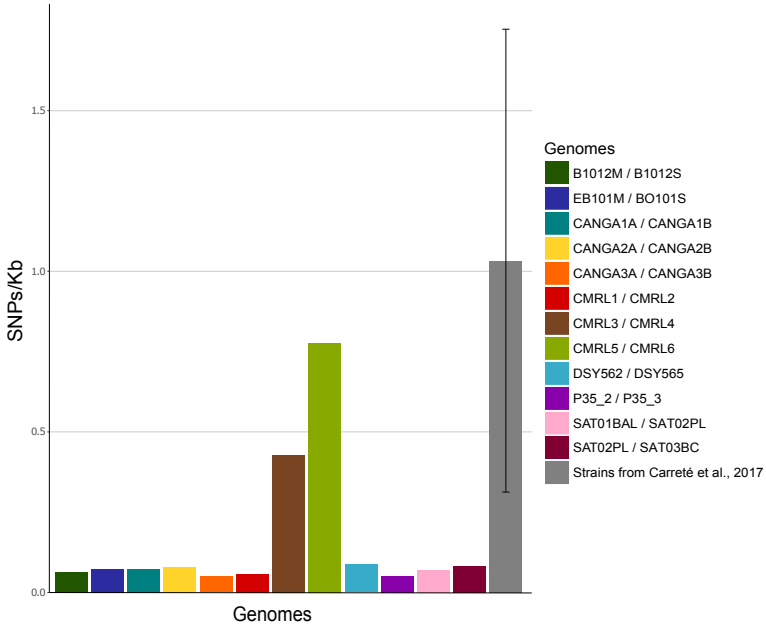

Figure 5.1: Amount of different SNPs/Kb per each pair of genomes. Grey bar indicates the average of all non-clonal strains from Carreté et al., 2017.

of SNPs found in two isolate pairs (CMRL5/CMRL6 and CMRL3/CMRL4) was significantly higher than the observed for the other sequence pairs, and within the range of SNP differences found between clinical isolates of different sources but within the same genetic clade (Figure 5.1). This shows evidence that strains in these two isolate pairs are actually non-clonal, and that they correspond to two divergent strains that belong to the same clade. This observation has gone unnoticed in previous analyses (Biswas et al., 2017).

Previously studies found that *C. glabrata* can undergo genomic rearrangements as an adaptive survival mechanism. To analyze the plasticity of these clonal *C. glabrata* genomes, we estimated the number of structural variants using depth of coverage analysis (See Material and Methods). Depth of coverage analysis revealed aneuploidies involving chromosome C and E in CMRL2 (Figure 5.2a), and thus these chromosomes were removed for the analysis of duplications. In addition, SAT strains displayed unusual higher coverage in subtelomeric regions, likely due to library preparation biases, and were not considered in the duplication analyses. Overall, in all analyzed strains we detect four deleted genes that were specific for the second isolated strain from the pair of samples (Figure 5.2b), three of them related with adhesin-like proteins (*EPA8*, CAGL0F09273g and CAGL0I11011g). Overall, we detect 46 genes that were specific for one of the two isolated strains in a pair.

For the nine clonal strain pairs we asked the question of whether common trends could be observed in terms of genes that present non-synonymous SNPs between the two strains of a pair. Overall, we detected a total of 187 genes affected by a non-synonymous mutation (Supplementary Table S1). Notably, in four pairs of strains we found fungal-type cell wall as an enriched term among genes having non-synonymous SNPs: CANGA1A/CANGA1B, CANGA2A/CANGA2B, CMRL1/CMRL2 and DSY562/DSY565. Remarkably, one gene CAGL0E00231g, encoding a putative adhesin-like protein, was found to present non-synonymous mutations in all analyzed pairs of strains. Such mutations affected different residues in the different strains.

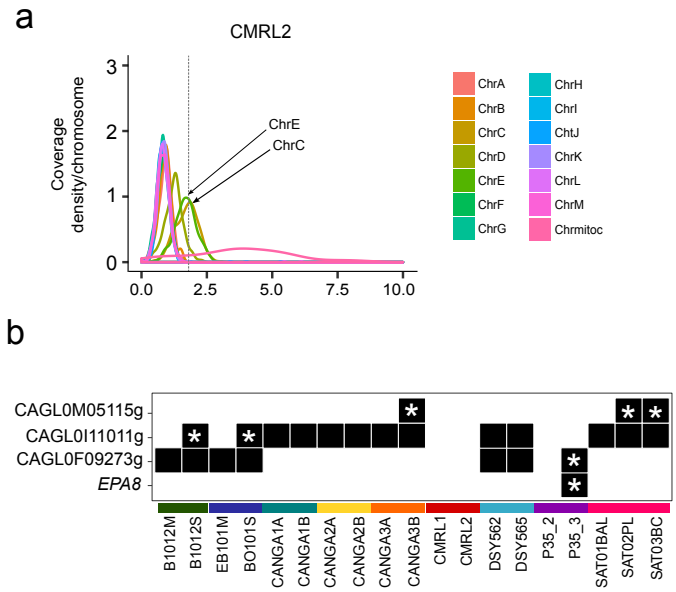

Figure 5.2: Structural variants in clonal isoaltes. a) Aneuploidy found in chromosomes C and E in CMRL2. Y-axis show the coverage density per chromosome. Vertical dashed line in X-axis indicates the threshold used to detect chromosome duplications (threshold 1.8). Chromosomes affected are marked with arrows. b) Heatmap showing deleted genes that appeared in the second isolates. Asterisk (∗) in boxes indicates second isolate affected by this deletion

Subsequently, we used pairwise comparisons of genomes from isolate pairs to analyze the distribution of SNP density using a non-overlapping 10Kb windows across the genome. Most strain pairs showed evenly distributed SNP densities across the entire genome (Supplementary Figure 5.6). Exceptionally, we observed that the identified mutations in the SAT strains were not evenly distributed along the genome, but clustered in specific, large regions of chromosomes L and M (Figure 5.3a). Of note, clade II (to which the SAT strains belong) has been previously described as a clade showing genetic admixture with clade I (Carreté et al., 2017). We thus investigated whether the observed patterns of SNPs in the SAT strains could be the result of genetic admixture with genetically distant strains.

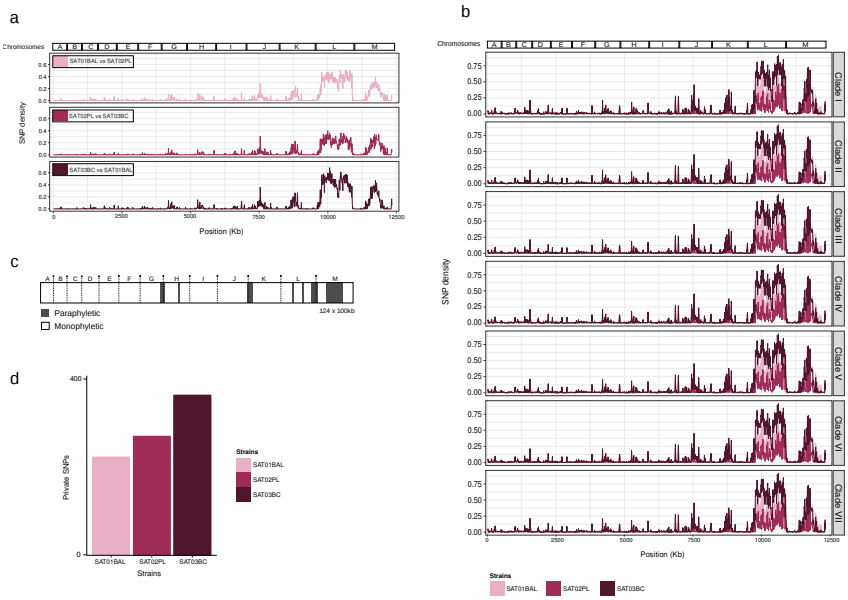

Figure 5.3: Pairwise comparison and SNP distribution across the three SAT isolates. a) SNP densities obtained comparing SAT strains using non-overlapping 10Kb window. First profile indicates SAT01BAL versus SAT02PL, second profile indicates SAT02PL versus SAT03BC, and third profile indicates SAT03BC versus SAT01BAL. Bar on the top indicates relative length and order of *C. glabrata* chromosomes. b) SNP densities obtained comparing SAT strains versus all available strains from around the globe (Carreté et al., 2017). Each profile shows the comparison between SAT strains and strains from each clade. c) Monophyletic (in white) and paraphyletic (in grey) regions across the SAT genome. d) Barplot showing the total number of SNPs that only appeared in one SAT strains.

5.3.2 Evidence for differential genetic admixture within the same host

Given the above described findings that most of the genetic differences between the SAT isolates concentrated in chromosomes L and M, we repeated the pairwise comparison with all the clades. By doing so, we expected to identify possible donor clades as those having few genetic differences with the putatively recombined regions in the SAT strains (Figure 5.3b). However, high density of SNPs in these regions from chromosomes L and M when compared with representatives of all tested

clades, suggesting that the putative donor strains does not belong to any of the presently sequenced clades. Moreover, as these recombined regions are markedly different between each of the three SAT strains, it also indicates that three different genetic backgrounds may have contributed, independently, to each of the SAT strains.

The only alternative scenario that could explain this finding would be that the putatively recombined regions are mutational hotspots, having a much larger mutation rate than the rest of the genome in these strains. We consider this explanation very unlikely, given the large differences in SNP densities (up to 0.68 SNP/Kb), and the observation that these SNPs maintained usual ratios of synonymous to non-synonymous SNPs. Nevertheless, we tested this possibility by performing phylogenetic analyses on genomic windows on those chromosomes, over an alignment containing the SAT strains and a representative set of 33 available strains from around the globe (Carreté et al., 2017) (See Material and Methods). We then reconstructed phylogenetic trees based on SNP data from non-overlapping 100Kb windows. If the putatively recombined regions were to be the result of accelerated evolutionary rates, phylogenetic trees from these regions should show a monophily of the SAT strains, and differences with trees from other genomic regions will be restricted to the differences in branch lengths. Contrary, we found that the putatively recombined regions largely overlap with windows producing paraphyletic relationships between SAT strains, were the expected monophily was broken by the presence of strains from other clades (Figure 5.3c). This finding reinforces the idea that these regions entail distinct evolutionary histories, which is compatible with recombination with distant clades.

These putatively recombined genomic areas correspond to roughly 94.48% of the chromosome L (137 highly different regions from 145 regions of 10Kb across the chromosome) and the 78.57% of the chromosome M (110 highly different regions from 140 regions of 10Kb across chromosome M). This observation brings about the possibility that mating between distant strains led to exchange of entire chromosomes, as it occurs in *C. albicans* parasexual cycle.

We obtained a list of single nucleotide polymorphisms that occur only in one of the pair of sequenced strains (i.e. private mutations). We observed that the number of private SNPs in increases with the time of isolation, with 226 SNPs in SAT01BAL, 274 SNPs in SAT02PL and 369 SNPs in SAT03BC (Supplementary Table S3, Figure 5.3d). However most (273/274) private SNPs in SAT02 PL had the same alternative nucleotide in the other two strain. Assuming the three strains represent subsequent isolates of an evolving clone this would imply a very high number of reversions, something we consider unlikely. Most parsimonious interpretation, particularly considering that the three isolations are separated by less than a week is that the three strains represent standing genetic variation within the host. Thus the availability of three serial strains, instead of two, allows us to differentiate between the contribution of standing variation and newly emerged mutations to the observed genetic differences between isolated strains.

5.3.3 Phenotypic analysis

Another ability of *C. glabrata* was the enormous plasticity in terms of genome variation such as SNPs or CNVs. Given the high variability that they have, one punctual genotypic change can drive to a phenotypic change. To asses whether the genomic plasticity affects the SAT strains at the phenotypic level, we measured antifungal drug susceptibility, biofilm formation and growth under seven different conditions. Those conditions were high and low pH, presence of DTT, high temperature (41.5°C), sodium chloride, hydrogen peroxide and also included the reference medium YPD as a control (see Materials and Methods). Most tested conditions resulted in a normal growth, with hydrogen peroxide and pH=2 as an exception (Figure 5.4a). In the presence of hydrogen peroxide, SAT03BC grows faster than SAT01BAL and SAT02PL; and SAT02PL grows faster than SAT01BAL and SAT03BC with pH=2. B1012M/B1012S, EB101M/BO101S and P35_2/P35_3 were analyzed previously in other study (Carreté et al., 2017), with apparent no phenotypic differences. In addition, we tested the resistance using different types of anti-fungals (amphoteticin B, fluconazole, voriconazole, posaconazole, isavuconazole, micafungin and caspofungin)

using EUCAST protocol (See Material and Methods). Our results did not show any differences related to drug resistance or sensitivity, in contrast to the observed results in two pairs of the strains used in this project (paper sanglard I paper biswas). We also analyzed differences on adherence properties. We observed significant differences between SAT01BAL and SAT03BC (*p*-value 0.0011 (t-student)) (Figure 5.4b). A specific non-synonymous mutation in SAT03BC was found in *SIR4*, a gene encoding a protein involved in subtelomeric silencing and regulation of biofilm formation (Iraqui et al., 2004). This genetic differences, together, with the enrichment in non-synonymous SNPs related to adhesins (see above), may underly the observed phenotypic variation in terms of adhesion properties in the SAT strains, although futur research is needed to confirm this.

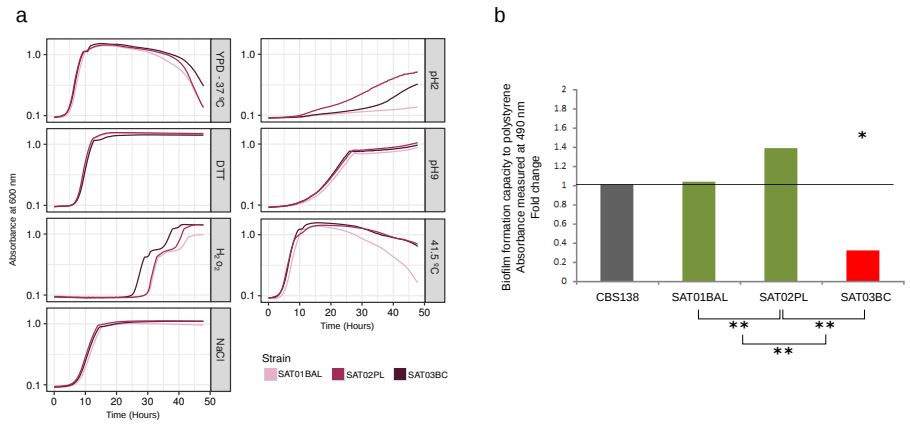

Figure 5.4: Phenotypic analysis of SAT strains. a) Growth curves for seven different conditions. YPD, H_2O_2, NaCl, DTT, high temperature (41.5 T°), basic pH (pH=9) and acid (pH=2). All curves were carreid out at 37 °C, with the exception of high temperature condition. X-axis shows time (in hours) and y-axis shows the OD value per each condition. b) Biofilms formation analysis. Barplots represent averages of three independent replicas of four technical repeats each. $*$ indicates if exist significantly differences between SAT01BAL-SAT02PL, SAT02PL-SAT03BC and SAT01BAL-SAT03BC (t – student, p-value $<$ 0.05). Significantly differences between reference CBS138 and SAT03BC p-value 0.0011 (barplot in coloured in red)

5.4 Conclusions

To obtain an overall view of the genomic variation of clinical samples inside a patient under treatment, we re-analyzed different samples from previous studies (Carreté et al., 2017; Håvelsrud and Gaustad, 2017; Biswas et al., 2017; Vale-Silva et al., 2017). In addition we sequenced a trio of serial isolates. The availability of three clinical samples instead of two, provides us with the possibility to disentangle *de novo* mutations from standing variation. Our results show the genomic variation that exists between clinical samples isolated from same patient. The presence of non-clonal strains in the same patient was detected in two out of the eleven analyzed pairs of strains. SAT strains present admixture between samples from clade I and clade II, previously described in Carreté et al. (2017). In addition, we detect recent admixture events that affect the three strains of the trio, all affecting the same two chromosomes but each strain recombining with a different genetic background. Interestingly these genetic backgrounds seem not to correspond to any of the known clades with sequenced representatives. We detect just one aneuploidy in eleven pairs of strains. Hence, recombination could be as common as aneuploidy in *C. glabrata*'s clinical isolates, although our sampling is admittedly restricted. Nevertheless, future studies including additional serial isolates will undoubtely help us to clarify how genomic variation and selection processes affect disease and treatment outcome.

5.5 Materials and Methods

Strains

The collection of 23 *Candida glabrata* strains used for the analyses in this study are listed in (Table 5.1). Three strains SAT01BAL (synonym EF54001Bal), SAT02PL (EF54001Per) and SAT03BC (EF54001Blo) were sequenced in this study (see below). They correspond to isolates from bronchiolo-alveolar lavage (BAL), peritoneal fluid (PL) and blood culture (BC), respectively, collected from the same patient. This 56 years old male patient was hospitalized in 2007 in a haematology department, in Paris area (France) for treatment of an acute myloblastic leukaemia. At the time of the firs sampling, he was have received antifungal prophylaxis with fluconazole for 2 weeks. The second and third samples were obtained one and six days, respectively, after the first sample.

Sequencing

The genome sequences for SAT01BAL, SAT02PL and SAT03BC strains were obtained at the Ultra-sequencing core facility of the CRG, using Illumina HiSeq2000 sequencing machines. Paired-end libraries were prepared. DNA was fragmented by nebulization or in Covaris to a size of $\tilde{6}$00 bp. After shearing, the ends of the DNA fragments were blunted with T4 DNA polymerase and Klenow fragment (New England Biolabs). DNA was purified with a QIAquick PCR purification kit (Qiagen). 3'-adenylation was performed by incubation with dATP and 3'-5'-exo- Klenow fragment (New England Biolabs). DNA was purified using MinElute spin columns (Qiagen) and double-stranded Illumina paired-end adapters were ligated to the DNA using rapid T4 DNA ligase (New England Biolabs). After another purification step, adapter-ligated fragments were enriched, and adapters were extended by selective amplification in an 18-cycle PCR reaction using Phusion DNA polymerase (Finnzymes). Libraries were quantified and loaded into Illumina flow-cells at concentrations of 7 - 20 pM. Cluster generation was performed in an Illumina cluster station. Sequence runs of 2x100 cycles were performed on the sequencing instrument. Base calling

was performed using Illumina pipeline software. In multiplexed libraries, we used 4 bp internal indexes (5 indexed sequences). De-convolution was performed using the CASAVA software (Illumina). Sequence data of the genomes has been deposited in SRA and will be available upon publication.

SNP calling

Reads were aligned onto the reference assembly of the CBS138 strain (Dujon et al., 2004) using BWA, with the BWA-MEM algorithm with 16 as number of threads (Li and Durbin, 2010). We identified SNPs using GATK v3.3 (McKenna et al., 2010; DePristo et al., 2011; Van der Auwera et al., 2013) with a haploid model, filtering out clusters of 5 variants within 20 bases and low quality variants, and using thresholds for mapping quality and read depth (>40 and >30 respectively).

Structural variants

We used deviation from the expected depth of coverage to detect structural variants (Boeva et al., 2011). For every *C. glabrata* strain we measured the total number of genes deleted and duplicated using depth of coverage analysis from Samtools (Li et al., 2009; Li, 2011). After mapping the reads to the CBS138 genome, a gene was considered deleted if less than 20% of the length of a given gene was covered by reads. For duplications and large scale structural variants, we normalized the number of reads per gene. We consider a duplication if the median coverage of a gene was 1.8 times or higher than the median coverage of the chromosome.

Recombination analysis

Pairwise comparison using SNP density with a non-overlapping 10Kb windows was used in order to detect visualy recombination between samples.

Phylogenetic analysis at chromosome level

By using the previously annotated SNPs, we reconstructed the sequence of each strain by replacing the reference nucleotide for a given SNP positions with coverage <20 were coded as gaps. Those genomes were aligned with reference strain CBS138 and the rest of *Candida glabrata* strains that were available (Carreté et al., 2017). Then, those genomes were divided using a non-overlapping 100Kb windows. We reconstructed a species tree per each fragment. The resulting alignment was trimmed using TrimAl v.1.4 (Capella-Gutiérrez et al., 2009) to delete positions with more than 50% gaps. Finally a phylogenetic tree was reconstructed from the trimmed alignment using RAxML v7.3.5, with model Protgammalg (Stamatakis et al., 2005).

Phenotypic analyses

Growth curves

SAT strains were recovered from our glycerol stock collection and grown for 2 days at 37 °C on a YPD agar plate. Single colonies were cultivated in 15 ml YPD medium (37 °C, 200 rpm, overnight). Then, each sample was diluted to an optical density (OD) at 600 nm of 0.2 in 3 ml of YPD medium and grown for 3 h more in the same conditions (37 °C, 200 rpm). Dilutions were made again to have an OD at 600 nm of 0.5 in 1 ml of YPD medium in order to have the same amount of cell in all the experiments. The samples were centrifuged for 2 min at 3000 g, washed with 1 ml of sterile water and centrifuged again for 2 min more at 3000 g for a final resuspension of the pellet in 1 ml of sterile water. At the end, 5 μl of each sample was inoculated in 95 μl of the corresponding medium in a 96-well plate. All experiments were run in triplicate.

We tested six different growth conditions: the oxidative stress was assessed by the growth of the cultures on YPD medium supplemented with 10 mM H_2O_2, reductive stress with 2.5 mM DTT and osmotic stress with 1 M NaCl, high temperature (41.5 °C), pH=2 and pH=9 along with the control growth on YPD itself. Cultures were grown in 96-well plates at 37 °C or 41.5 °C, shaking, for 24 or 72 h depending on the growth rate in each condition,

and monitored to determine the optical density at 600 nm every ten min by a TECAN Infinite®M200microplate reader. Growthcurver v0.2.1, an R package, was used to measured growth conditions (Sprouffske and Wagner, 2016).

Biofilm formation assay

Studied isolates and controls (CBS138, moderate biofilm formation capacity; PEU-382 and PEU-427, high biofilm formation capacity were cultured overnight in YPD medium at 37 °C. The optical density was determined at 600 nm (Ultrospec 1000) and adjusted to a value of 2 using sterile $NaCl_{physiol}$. 50 μl aliquots of the cell suspensions were placed into 96-well polystyrol microtiter plates (Greiner Bio-one) and incubated for 24 h at 37 °C. The medium was removed and the attached biofilms washed once with 200 l distilled water. Cells were stained for 30 min in 100 μl of 0.1% (w/v) crystal violet (CV) solution. Excess CV was removed and the biofilm carefully washed once with 200 μl distilled water. To release CV from the cells, 200 μl 1% (w/v) SDS in 50% (v/v) ethanol were added and the cellular material resuspended by pipetting. CV absorbance was quantified at 490 nm using a microtiter plate reader (MRX TC Revelation). Final data is the average of the three independent biological experiments, each one with four technical repeats.

Antifungal drug susceptibility testing

Isolates were cultured overnight on Sabouraud (Oxoid) agar plates. After that, antifungal drug susceptibilities towards Fluconazole, Isavuconazole, Posaconazole, Voriconazole, Micafungin, Caspofungin, 5-Fluorcytosine, and Amphotericin B were determined according to EUCAST EDef 7.1 method (Arendrup et al. 2012). The MIC values of each SAT strains were calculated according to EUCAST guidelines (http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/AFST/Clinical_breakpoints/Antifungal_breakpoints_v_8.0_November_2015.xlsx, accessed Nov 16th 2016)

Statistical analyses

Multiple Correspondence Analysis (MCA) was applied using ade4 package for R in roder to establish the main relationships between strains (Tenenhaus and Young, 1985).

5.6 Supplementary Figures and Tables

All Supplementary Tables are available in digital version. Supplementary Figures are in high quality also in digital version.

5.6.1 Supplementary Tables

Supplementary Table S1

List of genes affected by synonymous and non-synonymous mutations in clonal strains. Columns in the table indicates, in this order: Gene affected, standard name of that gene, description of gene affected.

Supplementary Table S2

Private mutations for SAT strains. Columns in the table indicate: chromosome and position affected by the SNP; Gene name; type of mutation (synonymous and non-synonymous); Amino acid substitution.

5.6.2 Supplementary Figures

Supplementary Figure 5.5

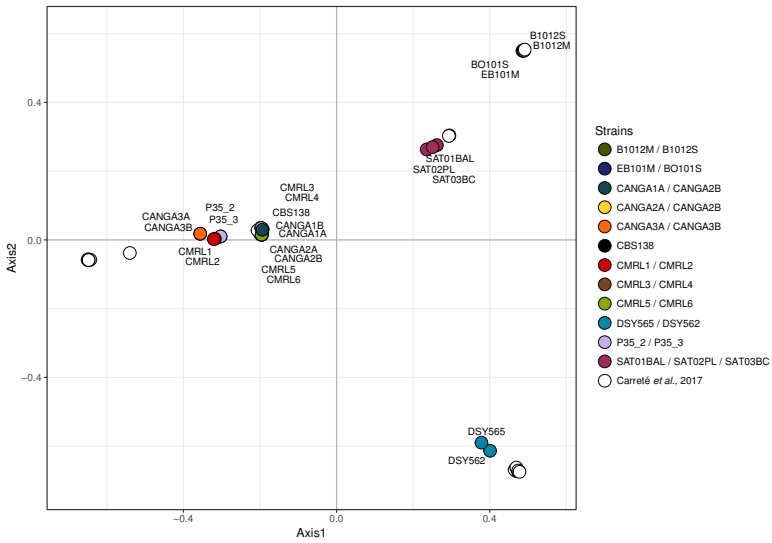

Figure 5.5: Distribution of strains based on SNPs using a Multiple Correspondence Analysis (MCA). Each pair of strains used marked in differents colors and named in plot. White circles in plot indicates strains used to provide a global comparison (Carreté et al., 2017).

Supplementary Figure 5.6

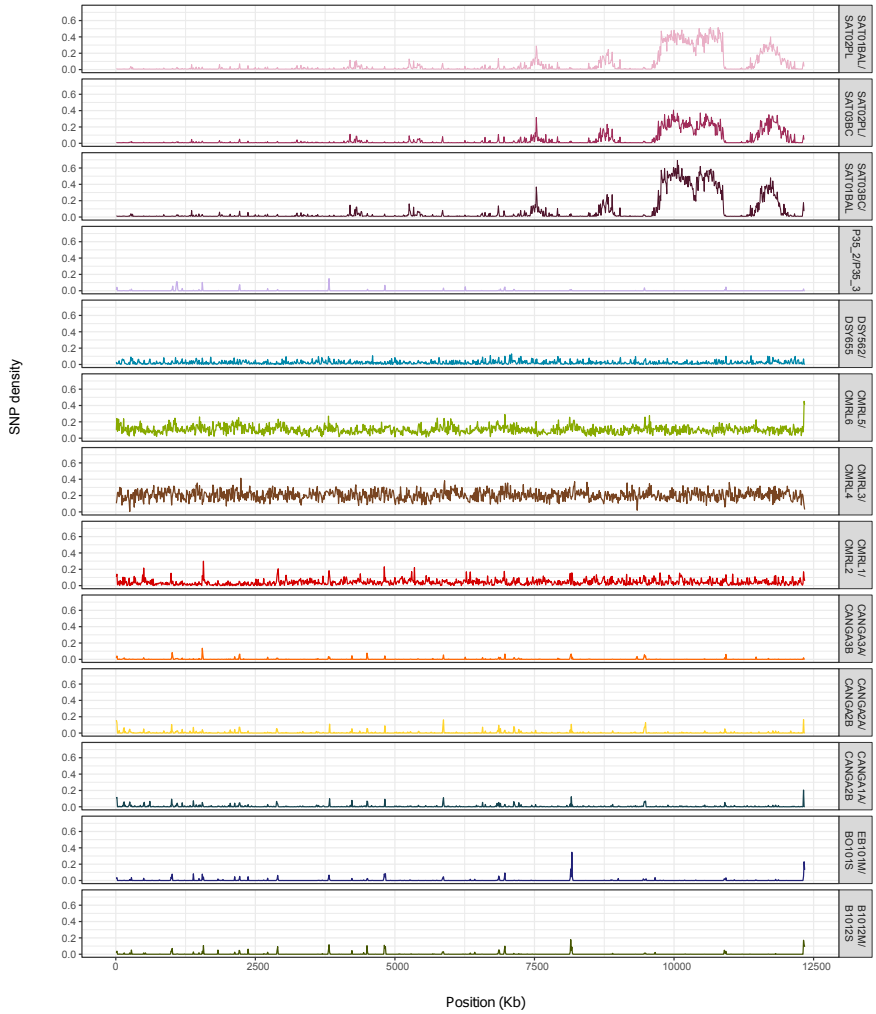

Figure 5.6: Distribution of SNP density usinf a non-overlapping 10Kb window across the genome. Each profile shows the pairwise comparisons of each pair of strains.

Part III

Discussion

6 Summarizing discussion

The overall aim of this thesis project was to gain insights into the recent evolution of the emerging pathogen *Candida glabrata*, through the use of comparative genomics and population genomics techniques. Central questions at the start of the project were the existence or not of a sexual cycle in *C. glabrata*, the search for possible co-evolutionary patterns that would indicate (or not) an ancient association with the human host, and the discovery of main genomic changes occurring in the global genetic landscape of this pathogen.

Candida glabrata at the front end of pathogenic yeasts

Fungal infections are becoming increasingly important during recent years, partly owing to medical progress, which promotes survival of susceptible individuals. The majority of fungal infections are caused by dermatophytes and any type of *Candida* yeasts. Candidiasis is the general term used to designate fungal infections caused by *Candida* yeasts. Invasive candidiasis - i.e. when yeasts invade tissues that are otherwise sterile- can range for 75% of all systemic fungal infections, and poses a serious threat to life, particularly in immunocompromised patients, with mortality rates reaching 46%-75% (Wilson et al., 2002). There are more than 30 *Candida* species that can cause Candidiasis, but the three most common species are *Candida albicans*, *Candida glabrata* and *Candida parapsilosis*, generally in this order (Diekema et al., 2012; Gabaldón et al., 2016).

Candida glabrata is more closely related to *Saccharomyces cerevisiae* than to *C. albicans* or *C. parapsilosis*. Most typical *Candida* pathogens belong to the *Candida* clade, a group of species that share an important characteristic in

their genetic code, in which the CUG codon encodes a Serine instead of a Leucine (Santos et al., 2011). However, *Candida glabrata* is an out-lier and belongs a different clade, called Saccharomycetacea. This clade contains *Saccharomyces cerevisiae*, the most extensive studied eukaryotic organism, and other species that are involved in industrial processes but where pathogens are not common. The fact that human fungal pathogens belong to different clades and that they all have close relatives that are non-pathogenic indicates that the capacity to infect humans has evolved several times independently, likely following different evolutionary routes.

The importance of identifying different evolutionary mechanisms that underly the emergence of virulence, and the associated genomic changes is of high importance. Understanding how pathogenesis evolves and appears during evolution can allow us to discover novel pathogenicity mechanism, and how emerging pathogens such as *C. glabrata* have adapted to the human host. Virulence in opportunistic pathogens can only be apparent under certain conditions. Pathogens may only express their virulence when they are in the wrong host or tissue or when they find a weakened host. This situation suggests that virulence in opportunistic pathogens is not an adaptive phenotype, but rather a secondary effect of adapting to some other selective pressure. An evolutionary accident rather than a goal in itself.

Generally, virulence results from an alteration of the host-microbe interactions. Given the same alteration on the host, some commensal species such as *C. albicans* and *C. glabrata* have a higher probability to cause disease than others, for reasons we do not fully understand. The availability of the genome sequence of *C. glabrata* (Dujon et al., 2004), opened the possibility of comparing it to the genomes of *C. albicans* (Jones et al., 2004) and *S. cerevisiae* (Goffeau et al., 1996). These comparisons revealed some clear differences. *C. glabrata* is a haploid organism instead of a diploid organism (such as *C. albicans*), it can survive within macrophages, can grow at 37 °C, it has high adherence to human tissues or clinical material and has the capacity to acquire drug resistance to commonly used drugs (see Table 3.1, Chapter 3). These abilities are useful for pathogenesis and a key to become an important pathogen.

C. glabrata may be adapted to non-human, environmental niches

C. glabrata belongs to a genus in Saccharomycetaceae family called *Nakaseomyces* (Kurtzman, 2003). Other species in this genus have an environmental lifestyle, with the exception of *Candida nivariensis* (Alcoba-Florez et al., 2005) and *Candida bracarensis* (Correia et al., 2006), both collected from human clinical samples. Besides clinical samples, *Candida nivariensis* have also been found in flowers and insects, suggesting that *C. nivariensis* can live associated to human tissues and another environments (Kurtzman et al., 2011). One *C. glabrata* strain isolated from fertile soil has been recently sequenced (Xu et al., 2016). Analyzing the SNP patterns of that *C. glabrata* strain we observed that it was remarkably close to the reference CBS138 genome (Carreté et al., 2017). Additionally, it is known that *C. glabrata* is a competent alcohol producing yeast and encodes six duplicated paralogs that encode glycolytic enzymes (Hagman et al., 2013). These findings may reflect adaptation to environmental niches (e.g. plants, fruits and/or insects) rather than the human body.

Virulence appeared multiple and independent times in the *C. glabrata* group

The genome sequencing of the close relatives of *C. glabrata* (Gabaldón et al., 2013) indicated that environmental species *N. bacillisporus* and *C. castellii* were very divergent from the pathogenic species. However, another non-pathogenic species, *Nakaseomyces delphensis*, was branching within the mildly pathogenic species *Candida nivariensis* and *C. bracarensis*, with *C. glabrata* as the first diverging lineage for this clade (Figure 3.1, Chapter 3). The origin of these three pathogens suggested that the virulence trait appeared multiple and independent times. Another observation was the high correlation between genes from the EPA family of adhesines and the ability to infect humans, suggesting an important role in virulence in this clade (Gabaldón et al., 2013).

High levels of genetic diversity between *Candida glabrata* clades

The above-mentioned results opened a door to a more detailed exploration of the evolution of *C. glabrata* at the species level. During the last year several studies have appeared that analyze genomes from clinical isolates from *C. glabrata* (Carreté et al., 2017; Håvelsrud and Gaustad, 2017; Biswas et al., 2017; Vale-Silva et al., 2017). The availability of diverse genomes helps to improve our understanding of the genomic variation in populations of this pathogen and how this relates to its epidemiology and pathogenecity. In the context of the present thesis up to 57 genomes were used to analyze in detail the genomic diversity across *C. glabrata* clinical isolates. Using different approaches focused on the comparison of SNP patterns we observed that the sequenced *C. glabrata* strains can be classified into seven distinct clades. Most clades comprise strains isolated from different locations around the globe and from different body sites, suggesting there is no correlation between genetic background, geography, and isolation site. We observed a low genetic variability in terms of SNPs between strains form the same clade, but high genetic distance between clades. Comparison between strains from the same patient presented very low genetic diversity, with a range of 0.05-0.07 SNPs/Kb. Recent studies based on sequential isolates show also the low variability between strains from same patient, for example one study just found 17 non-synonymous SNPs between isolates (Vale-Silva et al., 2017). The low variability between strains from same patient was indicative that patients were colonized by clonal strains, that could be distributed to different body sites. Unexpectedly, two pairs of strains from the same host presented non-clonal distribution, as indicated by their large genetic distance, similar to that found in independent isolates of the same clade.

Genome plasticity in *C. glabrata*

Structural variants can be detected using depth of coverage. Taking into account the 33 strains used in Chapter 4, we detected a total of 46 deletions and 62 duplicated genes. As mentioned above, adhesins are important to infect humans. Interestingly we found that genes encoding adhesin-like

proteins were enriched among genes deleted or duplicated, representing up to 45.65% and 41.94% of the cases, respectively. This enrichment suggest that there exists a standing variation of the adherence trait, which may be linked to an ongoing selection.

We found aneuploidies involving different chromosomes that affected genes related to drug resistance. Importantly, we found two events related to aneuploidies that underscore genome plasticity. First, we found two aneuploidies in the second isolate from a clinical series, never described before (Biswas et al., 2017), and a spontaneous aneuploidy while growing one strain under laboratory conditions. Both situations indicate a high genomic plasticity.

De novo assembly techniques helps to analyze large re-arrangements and investigate the core and the accessory material of a set of genomes. We assembled de novo the genomes for all strains and subsequently aligned them to the reference. We observed new large re-arrangements and confirmed previously described ones (Muller et al., 2009). The existence of re-arrangements confirms the high plasticity in *C. glabrata* and put de novo assembly as a useful complement to classical techniques such as PFGE (pulsed-field gel electrophoresis). In order to asses in more detail the plasticity in these strains, analysis of pan and accessory genome was performed. An average of 342 genes were unique for each strain, suggesting that some strain-specific genes were the result of, on the one hand, artifacts from the clustering approach, and on the other hand, newly emerged genes. The rapid availability of sequenced genomes together with the analysis of pan-genome, opens a door to discover the core and the accessory material that underlies the large phenotypic variability in *C. glabrata*.

Genomic recombination

Genomic admixture was found between strains from different clades (Clade I and Clade II). When these sequences were analyzed using pairwise comparisons of SNP density, large regions appeared without SNPs. In the case of an ancient variation we would expect that polymorphisms should

be spread throughout the entire genome, but the presence of these regions devoid of SNPs are indicative of recent recombination. Recombination was divided in recent and ancestral recombination. Here the term ancestral refers to events that predate the origin of each of the clades and is shared by all strains in a clade, and recent when it occurred after the origin of the defined clades and is restricted to one or few strains within a clade. In general, recent recombination was still playing a role in all chromosomes, but ancestral recombination was more common than recent recombination. An example of clear recombination was found in three strains from Clade II (see Chapter 5). These three strains from Clade II presented a first event of recombination with Clade I and Clade II, and a second and recent recombination with other external *C. glabrata* strain. In contrast, the rest of strains from Clade II present just the first recombination with Clade I.

Recombination in the *C. glabrata* population was briefly described before using specific genes or using multilocus sequence typing (MLST) (Cormack and Falkow, 1999; Dodgson et al., 2005). Thanks to the availability of full genomes sequences we can test the presence of recombination events across the genome. Recombination could be explained by deletions or duplications shared between different lineages. The presence of recombination hotspots across the sequences of genes affected by CNV or the clusterization of strains from different clades using phylogenetic networks, suggested that these structural variants were the result of genetic exchange mediated by recombination. We conclude that *C. glabrata* can recombine inter-clade and intra-clade and that processes of genetic admixture are ongoing.

Mating type switching

The recent discovery of the possibility that distinct lineages of *C. glabrata* can recombine, implies necessarily the existence of mating. Mating type switching is the process that allows changing from one mating type, **a** or alpha, to another, using a complex mechanism that involves recombination between different genomic loci. Previous efforts have tried to prove the existence of mating type switching in *C. glabrata*. But despite great efforts,

switching was not observed in natural conditions. Only when switching is induced artificially by expressing the *S. cerevisiae HO* endonuclease it can be observed but it results in a high mortality (Boisnard et al., 2015). Nonetheless, the fact that *C. glabrata* genomes encodes all genes necessary to perform mating type switching, point to the possibility that mating-type switching can occur naturally under unknown conditions.

Our data supports the existence of mating-type switching in eight strains. In three cases we observed a normal switching from **a** to alpha. The five remaining, presented cases of aberrant switching, resulting from illegitimate gene conversions. In all aberrant cases, the correlation of the boundaries with the *HO* endonuclease cutting site, suggested that these switching events were the result of an illegitimate cuts in *MTL2* and *MTL3*. Additionally, the existence of some form of sexual cycle is supported by the signatures of purifying selection at the species level. Assessing levels of nucleotide diversity at synonymous and non-synonymous polymorphisms in *C. glabrata* genes, and compared these to *C. albicans* and *S. cerevisiae*, we observed similar levels of selective constraints. We observed that genes involved in meiotic recombination and repair had signs of relaxation of selection. In contrast, genes involved in mating or other processes of sexual cycle were more constrained.

Phenotypic differences between and within clades

Adherence properties can vary across strains because they are related to the genetic repertoire of encoded proteins that are attached to the cell-wall. Our results of copy number variation analysis shows an enrichment of adhesin-like proteins, one of the principal components of the cell-wall involved in adherence to various surfaces. The expansion of genes coding for adhesine-like proteins has been proposed a key in the emergence of the ability to infect human tissues in *C. glabrata* (Gabaldón et al., 2013). Adherence analysis revealed three strains with high ability to form biofilms. These strains shared independent duplications of *PWP4* and deletions of *AWP13*. Both genes are related to GPI-anchored adhesins (de Groot et al.,

2013), supporting the importance of adhesin proteins in biofilm formation. Biofilms can enhance resistance to antifungals and the host immune system, increasing the capacity to persist in the host. Acquiring resistance to drugs is an important property in *C. glabrata*. For this, testing drug resistance is an important issue. Three different strains showed a reduction of sensitivity to one or more antifungal drugs. Each strain carried a different private mutation in *PDR1*, a regulator of pleiotropic drug response (Tsai et al., 2006). Additionally, 45% of the analyzed strains carry a non-synonymous mutation in the *MSH2* gene (Healey et al., 2016). *MSH2* is involved in mismatch repair, and mutations in this gene have been proposed to confer a mutator phenotype that promotes the appearance of drug resistance (Healey et al., 2016). However, in our dataset, mutated *MSH2* variants were shared by all strains in the same clade (Clade II, III and IV) but without any unusual pattern of synonymous or non-synonymous SNPs. From this we conclude that these mutations represent natural genetic variation and do not confer a mutator phenotype.

Structural variants at the genomic level can affect the phenotype of the corresponding strain. We analyzed the growth of all sequenced isolates under different stress conditions. Most conditions showed differences, even between strains of the same clade. Altogether the large observed phenotypic variation, likely related with the underlying genomic plasticity, make knowledge of the clade affiliation a poor predictor of the behavior under the tested conditions. However, this type of analysis paves the way for future associations between specific genetic variations and phenotypes. In order to achieve this, a significantly expanded dataset of strains with available genomes and phenotypes will be needed.

7 Conclusions

- According to differences in single nucleotide polymorphisms, the sequenced *Candida glabrata* strains can be clustered into seven different clades. These clades show large genetic distances among them, suggesting they correspond to deeply diverged sub-populations.
- Lack of geographical structure of the observed genetic diversity, suggest recent global expansion of the clades.
- Genomic re-arrangements, copy number variations, and aneuploidies, contribute to genetic differences, even between closely related strains of the same clade. These are reflected into sometimes large phenotypic differences between related strains. As a result, clade assignment is a poor predictor of phenotypic behavior.
- Genomic variation is significantly enriched in cell-wall proteins. The highly dynamic repertoire of adhesines and the strain-specific differences of it, indicate the presence of standing variation of this emergent trait with an on going selection.
- Mating-type switching can occur in *C. glabrata* populations, however, there is a high degree of illegitimate recombination between mating-type loci.
- We show first genomic evidence of recombination and genetic admixture between distantly related strains.
- Genes involved in mating are evolutionarily constrained in *C. glabrata*. This, together with the existence of genetic admixture and recombination, shows evidence for the mating ability of *C. glabrata*, and the presence of a sexual or parasexual cycle.

- Most candidiasis patients included in our study are colonized by a clonal population of *C. glabrata*. However, the amount of standing variation can be high. Two of the patients were found to be colonized each by at least two genetically distant strains.

- Genetic diversity present within a patient can include strains showing evidence for distinct recombination events with distant clades, and with different phenotypes. Suggesting genetic admixture between co-infecting or co-colonizing *C. glabrata* strains may play a role in infection and clinical outcome.

8 Appendix: List of publications

1. Toni Gabaldón & **Laia Carreté**. (2016) The bird of a deadly yeast: tracing the evolutionary emergence of virulence traits in *Candida glabrata*. *FEMS Yeast Research*.

2. **Carreté L**, Ksiezopolska E, Pegueroles C, Gómez-Molero E, Saus E, Iraola-Guzmán S, Loska D, Bader O, Fairhead C, and Gabaldón T. (2017) Patterns of genomic variation in the opportunistic pathogen *Candida glabrata* suggest the existence of mating and a secondary association to the human host. *bioRxiv*.

3. **Carreté L**, Ksiezopolska E, Gómez-Molero E, and Gabaldón T. (2017) Genomic analysis of serial clinical isolates of *C. glabrata* suggests that mating and genetic introgression can occur during the course of infection. *(In preparation)*

4. **Carreté L**, Pegueroles C, and Gabaldón T. Population genomics in *Candida albicans*. *(In preparation)*

5. Figueras A, Robledo D, Corvelo A, Hermida M, Pereiro P, Rubiolo JA, Gómez-Garrido J, **Carreté L**, Bello X, Gut M, Gut IG, Marcet-Houben M, Forn-Cun G, Galán B, García JL, Abal-Fabeiro JL, Pardo BG, Taboada X, Fernández C, Vlasova A, Hermoso-Pulido A, Guigó R, Álvarez-Dios JA, Gómez-Tato A, Viñas A, Maside X, Gabaldón T, Novoa B, Bouza C, Alioto T, Martínez P. (2016) Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): a fish adapted to demersal life. *DNA research* Jun;23(3):181-92

References

Alcoba-Florez, J., Mendez-Alvarez, S., Cano, J., Guarro, J., Perez-Roth, E., and del Pilar Arevalo, M. (2005). Phenotypic and Molecular Characterization of Candida nivariensis sp. nov., a Possible New Opportunistic Fungus. *Journal of Clinical Microbiology*, 43(8):4107–4111.

Anderson, H. W. (1917). Yeast-Like Fungi of the Human Intestinal Tract. *Journal of Infectious Diseases*, 21(4):341–354.

Andersson, S. G. and Kurland, C. G. (1998). Reductive evolution of resident genomes. *Trends in microbiology*, 6(7):263–8.

Andrulis, E. D., Zappulla, D. C., Ansari, A., Perrod, S., Laiosa, C. V., Gartenberg, M. R., and Sternglanz, R. (2002). Esc1, a nuclear periphery protein required for Sir4-based plasmid anchoring and partitioning. *Molecular and cellular biology*, 22(23):8292–301.

Angiuoli, S. V. and Salzberg, S. L. (2011). Mugsy: fast multiple alignment of closely related whole genomes. *Bioinformatics (Oxford, England)*, 27(3):334–342.

Asai, K., Tsuchimori, N., Okonogi, K., Perfect, J. R., Gotoh, O., and Yoshida, Y. (1999). Formation of azole-resistant Candida albicans by mutation of sterol 14-demethylase P450. *Antimicrobial agents and chemotherapy*, 43(5):1163–9.

Auton, A. and McVean, G. (2007). Recombination rate estimation in the presence of hotspots. *Genome research*, 17(8):1219–27.

Bader, O., Schwarz, A., Kraneveld, E. A., Tangwattanachuleeporn, M., Tangwattanchuleeporn, M., Schmidt, P., Jacobsen, M. D., Gross, U., De Groot, P. W. J., and Weig, M. (2012). Gross karyotypic and phenotypic alterations among different progenies of the Candida glabrata CBS138/ATCC2001 reference strain. *PloS one*, 7(12):e52218.

Bennett, R. J. (2015). The parasexual lifestyle of Candida albicans. *Current opinion in microbiology*, 28:10–17.

Bennett, R. J., Forche, A., and Berman, J. (2014). Rapid Mechanisms for Generating Genome Diversity: Whole Ploidy Shifts, Aneuploidy, and Loss of Heterozygosity. *Cold Spring Harbor Perspectives in Medicine*, 4(10):a019604–a019604.

Bishop, J. A., Chase, N., Magill, S. S., Kurtzman, C. P., Fiandaca, M. J., and Merz, W. G. (2008). Candida bracarensis Detected among Isolates of Candida glabrata by Peptide Nucleic Acid Fluorescence In Situ Hybridization: Susceptibility Data and Documentation of Presumed Infection. *Journal of Clinical Microbiology*, 46(2):443–446.

Biswas, C., Chen, S.-A., Halliday, C., Kennedy, K., Playford, E., Marriott, D., Slavin, M., Sorrell, T., and Sintchenko, V. (2017). Identification of genetic markers of resistance to echinocandins, azoles and 5-fluorocytosine in Candida glabrata by next-generation sequencing: a feasibility study. *Clinical Microbiology and Infection*.

Bleidorn, C. (2017). *Phylogenomics*. Springer International Publishing, Cham.

Boeva, V., Zinovyev, A., Bleakley, K., Vert, J.-P., Janoueix-Lerosey, I., Delattre, O., and Barillot, E. (2011). Control-free calling of copy number alterations in deep-sequencing data using GC-content normalization. *Bioinformatics (Oxford, England)*, 27(2):268–9.

Boisnard, S., Li, Y. Z., Arnaise, S., Sequeira, G., Raffoux, X., Enache-Angoulvant, A., Bolotin-Fukuhara, M., and Fairhead, C. (2015). Efficient mating-Type switching in candida glabrata induces cell death. *PLoS ONE*, 10(10):1–18.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120.

Bolotin-Fukuhara, M. and Fairhead, C. (2014). Candida glabrata : a deadly companion? *Yeast*, 31(8):279–288.

Brisse, S., Pannier, C., Angoulvant, A., De Meeus, T., Diancourt, L., Faure, O., Muller, H., Peman, J., Viviani, M. A., Grillot, R., Dujon, B., Fairhead, C., and Hennequin, C. (2009). Uneven distribution of mating types among genotypes of Candida glabrata isolates from clinical samples. *Eukaryotic Cell*, 8(3):287–295.

Brown, G. D., Denning, D. W., Gow, N. A. R., Levitz, S. M., Netea, M. G., and White, T. C. (2012a). Hidden Killers: Human Fungal Infections. *Science Translational Medicine*, 4(165rv13):1–9.

Brown, G. D., Denning, D. W., Gow, N. A. R., Levitz, S. M., Netea, M. G., and White, T. C. (2012b). Hidden Killers: Human Fungal Infections. *Science Translational Medicine*, 4(165rv13):1–9.

Brunke, S. and Hube, B. (2013). Two unlike cousins: Candida albicans and C.glabrata infection strategies. *Cellular Microbiology*, 15(5):701–708.

Butler, G., Kenny, C., Fagan, A., Kurischko, C., Gaillardin, C., and Wolfe, K. H. (2004). Evolution of the MAT locus and its Ho endonuclease in yeast species. *Proc Natl Acad Sci U S A*, 101(6):1632–1637.

Butler, G., Rasmussen, M. D., Lin, M. F., Santos, M. A. S., Sakthikumar, S., Munro, C. A., Rheinbay, E., Grabherr, M., Forche, A., Reedy, J. L., Agrafioti, I., Arnaud, M. B., Bates, S., Brown, A. J. P., Brunke, S., Costanzo, M. C., Fitzpatrick, D. A., de Groot, P. W. J., Harris, D., Hoyer, L. L., Hube, B., Klis, F. M., Kodira, C., Lennard, N., Logue, M. E., Martin, R., Neiman, A. M., Nikolaou, E., Quail, M. A., Quinn, J., Santos, M. C., Schmitzberger, F. F., Sherlock, G., Shah, P., Silverstein, K. A. T., Skrzypek, M. S., Soll, D., Staggs, R., Stansfield, I., Stumpf, M. P. H., Sudbery, P. E., Srikantha, T., Zeng, Q., Berman, J., Berriman, M., Heitman, J., Gow, N. A. R., Lorenz, M. C., Birren, B. W., Kellis, M., and Cuomo, C. A. (2009). Evolution of pathogenicity and sexual reproduction in eight Candida genomes. *Nature*, 459(7247):657–662.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1):421.

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics (Oxford, England)*, 25(15):1972–3.

Capriotti, A. (1961). Torulopsis castellii sp.nov., a Yeast Isolated from a Finnish Soil. *Journal of General Microbiology*, 26(1):41–43.

Carreté, L., Ksiezopolska, E., Pequeroles, C., Gómez-Molero, E., Saus, E., Iraola-Guzmán, S., Loska, D., Bader, O., Fairhead, C., and Gabaldón, T. (2017). Patterns of genomic variation in the opportunistic pathogen Candida glabrata suggest the existence of mating and a secondary association to the human host. *bioRxiv*.

Casadevall, A., Fang, F. C., and Pirofski, L.-a. (2011). Microbial Virulence as an Emergent Property: Consequences and Opportunities. *PLoS Pathogens*, 7(7):e1002136.

Charlesworth, B. (2010). Molecular population genomics: a short history. *Genetics Research*, 92(5-6):397–411.

Cleveland, A. A., Farley, M. M., Harrison, L. H., Stein, B., Hollick, R., Lockhart, S. R., Magill, S. S., Derado, G., Park, B. J., and Chiller, T. M. (2012). Changes in Incidence and Antifungal Drug Resistance in Candidemia: Results From Population-Based Laboratory Surveillance in Atlanta and Baltimore, 2008-2011. *Clinical Infectious Diseases*, 55(10):1352–1361.

Conrad, D. F. and Hurles, M. E. (2007). The population genetics of structural variation. *Nature Genetics*, 39(7s):S30–S36.

Cormack, B. P. and Falkow, S. (1999). Efficient Homologous and Illegitimate Recombination in the Opportunistic Yeast Pathogen Candida glabrata. *Genetics*, 151(3).

Corradi, N. and Slamovits, C. H. (2011). The intriguing nature of microsporidian genomes. *Briefings in Functional Genomics*, 10(3):115–124.

Correia, A., Sampaio, P., James, S., and Pais, C. (2006). Candida bracarensis sp. nov., a novel anamorphic yeast species phenotypically similar to Candida glabrata. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, 56(1):313–317.

Coste, A., Turner, V., Ischer, F., Morschhäuser, J., Forche, A., Selmecki, A., Berman, J., Bille, J., and Sanglard, D. (2005). 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*, 172(4):2139–2156.

Cromie, G. a., Hyma, K. E., Ludlow, C. L., Garmendia-Torres, C., Gilbert, T. L., May, P., Huang, A. a., Dudley, A. M., and Fay, J. C. (2013). Genomic sequence diversity and population structure of Saccharomyces cerevisiae assessed by RAD-seq. *G3 (Bethesda, Md.)*, 3(12):2163–71.

Cui, L., Morris, A., and Ghedin, E. (2013). The human mycobiome in health and disease. *Genome Medicine*, 5(7):63.

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research*, 14(7):1394–403.

Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one*, 5(6):e11147.

Darwin, C. (1859). The origin of species by means of natural selection. *London: Murray Google Scholar*.

de Groot, P. W. J., Bader, O., de Boer, A. D., Weig, M., and Chauhan, N. (2013). Adhesins in human fungal pathogens: glue with plenty of stick. *Eukaryotic cell*, 12(4):470–81.

de Meeûs, T., Renaud, F., Mouveroux, E., Reynes, J., Galeazzi, G., Mallié, M., and Bastide, J. M. (2002). Genetic structure of Candida glabrata populations in AIDS and non-AIDS patients. *Journal of clinical microbiology*, 40(6):2199–206.

Delsuc, F., Brinkmann, H., and Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, 6(5):361–375.

Demuyser, L., Swinnen, E., Fiori, A., Herrera-Malaver, B., Vestrepen, K., and Van Dijck, P. (2017). Mitochondrial Cochaperone Mge1 Is Involved in Regulating Susceptibility to Fluconazole in Saccharomyces cerevisiae and Candida Species. *mBio*, 8(4):e00201–17.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., and Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, 43(5):491–8.

Diderrich, R., Kock, M., Maestre-Reyna, M., Keller, P., Steuber, H., Rupp, S., Essen, L.-O., and Mösch, H.-U. (2015). Structural Hot Spots Determine Functional Diversity of the Candida glabrata Epithelial Adhesin Family. *Journal of Biological Chemistry*, 290(32):19597–19613.

Diekema, D., Arbefeville, S., Boyken, L., Kroeger, J., and Pfaller, M. (2012). The changing epidemiology of healthcare-associated candidemia over three decades. *Diagnostic microbiology and infectious disease*, 73(1):45–8.

Dietrich, F. S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pöhlmann, R., Luedi, P., Choi, S., Wing, R. A., Flavier, A., Gaffney, T. D., and Philippsen, P. (2004). The Ashbya gossypii Genome as a Tool for Mapping the Ancient Saccharomyces cerevisiae Genome. *Science*, 304(5668):304–307.

Dodgson, A. R., Pujol, C., Denning, D. W., Soll, D. R., and Fox, A. J. (2003). Multilocus sequence typing of Candida glabrata reveals geographically enriched clades. *Journal of clinical microbiology*, 41(12):5709–17.

Dodgson, A. R., Pujol, C., Pfaller, M. A., Denning, D. W., and Soll, D. R. (2005). Evidence for recombination in Candida glabrata. *Fungal Genetics and Biology*, 42(3):233–243.

Domergue, R., Castaño, I., De Las Peñas, A., Zupancic, M., Lockatell, V., Hebel, J. R., Johnson, D., and Cormack, B. P. (2005). Nicotinic Acid Limitation Regulates Silencing of Candida Adhesins During UTI. *Science*, 308(5723):866–870.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuvéglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.-M., Beyne, E., Bleykasten, C., Boisramé, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.-M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.-F., Straub, M.-L., Suleau, A., Swennen, D., Tekaia, F., Wésolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P., and Souciet, J.-L. (2004). Genome evolution in yeasts. *Nature*, 430(6995):35–44.

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

Dunkel, N., Blaß, J., Rogers, P. D., and Morschhäuser, J. (2008). Mutations in the multi-drug resistance regulator MRR1, followed by loss of heterozygosity, are the main cause of MDR1 overexpression in fluconazole-resistant Candida albicans strains. *Molecular Microbiology*, 69(4):827–840.

Ehrlich, G. D., Hiller, N. L., and Hu, F. (2008). What makes pathogens pathogenic. *Genome Biology*, 9(6):225.

Eisen, J. A. and Fraser, C. M. (2003). Phylogenomics: Intersection of Evolution and Genomics. *Science*, 300(5626):1706–1707.

Enache-Angoulvant, A., Bourget, M., Brisse, S., Stockman-Pannier, C., Diancourt, L., François, N., Rimek, D., Fairhead, C., Poulain, D., and Hennequin, C. (2010). Multilocus microsatellite markers for molecular typing of Candida glabrata: application to analysis of genetic relationships between bloodstream and digestive system isolates. *Journal of clinical microbiology*, 48(11):4028–34.

Fabre, E., Muller, H., Therizols, P., Lafontaine, I., Dujon, B., and Fairhead, C. (2005). Comparative genomics in hemiascomycete yeasts: Evolution of sex, silencing, and subtelomeres. *Molecular Biology and Evolution*, 22(4):856–873.

Falush, D., Stephens, M., and Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4):1567–87.

Falush, D., Stephens, M., and Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular ecology notes*, 7(4):574–578.

Fanning, S. and Mitchell, A. P. (2012). Fungal Biofilms. *PLoS Pathogens*, 8(4):e1002585.

Fidel, P. L., Vazquez, J. A., and Sobel, J. D. (1999). Candida glabrata: review of epidemiology, pathogenesis, and clinical disease with comparison to C. albicans. *Clinical microbiology reviews*, 12(1):80–96.

Fischer, S., Brunk, B. P., Chen, F., Gao, X., Harb, O. S., Iodice, J. B., Shanmugam, D., Roos, D. S., and Stoeckert, C. J. (2011). Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes Into New Ortholog Groups. In *Current Protocols in Bioinformatics*, volume Chapter 6, pages Unit 6.12.1–19. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Fitzpatrick, D. A., Logue, M. E., Stajich, J. E., and Butler, G. (2006). A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evolutionary Biology*, 6(1):99.

Forche, A., Alby, K., Schaefer, D., Johnson, A. D., Berman, J., and Bennett, R. J. (2008). The parasexual cycle in Candida albicans provides an alternative pathway to meiosis for the formation of recombinant strains. *PLoS biology*, 6(5):e110.

Ford, C. B., Funt, J. M., Abbey, D., Issi, L., Guiducci, C., Martinez, D. A., Delorey, T., Li, B. Y. u., White, T. C., Cuomo, C., Rao, R. P., Berman, J., Thompson, D. A., and Regev, A. (2015). The evolution of drug resistance in clinical isolates of Candida albicans. *eLife*, 4:e00662.

Fundyga, R. E., Lott, T. J., and Arnold, J. (2002). Population structure of Candida albicans, a member of the human flora, as determined by microsatellite loci. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 2(1):57–68.

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

Gabaldón, T. and Carreté, L. (2016). The birth of a deadly yeast: tracing the evolutionary emergence of virulence traits in Candida glabrata. *FEMS Yeast Research*, 16(2):fov110.

Gabaldón, T., Martin, T., Marcet-Houben, M., Durrens, P., Bolotin-Fukuhara, M., Lespinet, O., Arnaise, S., Boisnard, S., Aguileta, G., Atanasova, R., Bouchier, C., Couloux, A., Creno, S., Almeida Cruz, J., Devillers, H., Enache-Angoulvant, A., Guitard, J., Jaouen, L., Ma, L., Marck, C., Neuvéglise, C., Pelletier, E., Pinard, A., Poulain, J., Recoquillay, J., Westhof, E., Wincker, P., Dujon, B., Hennequin, C., and Fairhead, C. (2013). Comparative genomics of emerging pathogens in the Candida glabrata clade. *BMC genomics*, 14:623.

Gabaldón, T., Naranjo-Ortíz, M. A., and Marcet-Houben, M. (2016). Evolutionary genomics of yeast pathogens in the Saccharomycotina. *FEMS Yeast Research*, 16(6):fow064.

Gasperskaja, E. and Kučinskas, V. (2017). The most common technologies and tools for functional genome analysis. *Acta medica Lituanica*, 24(1):1–11.

Ghannoum, M. A., Jurevic, R. J., Mukherjee, P. K., Cui, F., Sikaroodi, M., Naqvi, A., and Gillevet, P. M. (2010). Characterization of the Oral Fungal Microbiome (Mycobiome) in Healthy Individuals. *PLoS Pathogens*, 6(1):e1000713.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. G. (1996). Life with 6000 genes. *Science (New York, N.Y.)*, 274(5287):546, 563–7.

Gómez-Molero, E., de Boer, A. D., Dekker, H. L., Moreno-Martínez, A., Kraneveld, E. A., Ichsan, Chauhan, N., Weig, M., de Soet, J. J., de Koster, C. G., Bader, O., and de Groot, P. W. J. (2015). Proteomic analysis of hyperadhesive Candida glabrata clinical isolates reveals a core wall proteome and differential incorporation of adhesins. *FEMS yeast research*, 15(8):fov098.

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., and Shabalov, I. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research*, 42(D1):D699–D704.

Grimley, P. M., Wright, L. D., and Jennings, A. E. (1965). TORULOPSIS GLABRATA INFECTION IN MAN. *American journal of clinical pathology*, 43:216–23.

Haber, J. E. (2012). Mating-Type Genes and MAT Switching in Saccharomyces cerevisiae. *Genetics*, 191(1):33–64.

Hagman, A., Säll, T., Compagno, C., and Piskur, J. (2013). Yeast Make-Accumulate-Consume Life Strategy Evolved as a Multi-Step Process That Predates the Whole Genome Duplication. *PLoS ONE*, 8(7):e68734.

Hanson, S. J., Byrne, K. P., and Wolfe, K. H. (2014). Mating-type switching by chromosomal inversion in methylotrophic yeasts suggests an origin for the three-locus Saccharomyces cerevisiae system. *Proceedings of the National Academy of Sciences of the United States of America*, 111(45):E4851–8.

Hanson, S. J. and Wolfe, K. H. (2017). An Evolutionary Perspective on Yeast Mating-Type Switching. *Genetics*, 206(1):9–32.

Håvelsrud, O. E. and Gaustad, P. (2017). Draft Genome Sequences of Candida glabrata Isolates 1A, 1B, 2A, 2B, 3A, and 3B. *Genome announcements*, 5(10).

Hazen, K. C. (1995). New and emerging yeast pathogens. *Clinical microbiology reviews*, 8(4):462–78.

Healey, K. R., Zhao, Y., Perez, W. B., Lockhart, S. R., Sobel, J. D., Farmakiotis, D., Kontoyiannis, D. P., Sanglard, D., Taj-Aldeen, S. J., Alexander, B. D., Jimenez-Ortigosa, C., Shor, E., and Perlin, D. S. (2016). Prevalent mutator genotype identified in fungal pathogen Candida glabrata promotes multi-drug resistance. *Nature Communications*, 7:11128.

Hibbett, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O. E., Huhndorf, S., James, T., Kirk, P. M., Lücking, R., Thorsten Lumbsch, H., Lutzoni, F., Matheny, P. B., McLaughlin, D. J., Powell, M. J., Redhead, S., Schoch, C. L., Spatafora, J. W., Stalpers, J. A., Vilgalys, R., Aime, M. C., Aptroot, A., Bauer, R., Begerow, D., Benny, G. L., Castlebury, L. A., Crous, P. W., Dai, Y.-C., Gams, W., Geiser, D. M., Griffith, G. W., Gueidan, C., Hawksworth, D. L., Hestmark, G., Hosaka, K., Humber, R. A., Hyde, K. D., Ironside, J. E., Kõljalg, U., Kurtzman, C. P., Larsson, K.-H., Lichtwardt, R., Longcore, J., Miądlikowska, J., Miller, A., Moncalvo, J.-M., Mozley-Standridge, S., Oberwinkler, F., Parmasto, E., Reeb, V., Rogers, J. D., Roux, C., Ryvarden, L., Sampaio, J. P., Schüßler, A., Sugiyama, J., Thorn, R. G., Tibell, L., Untereiner, W. A., Walker, C., Wang, Z., Weir, A., Weiss, M., White, M. M., Winka, K., Yao, Y.-J., and Zhang, N. (2007). A higher-level phylogenetic classification of the Fungi. *Mycological Research*, 111(5):509–547.

Hirakawa, M. P., Martinez, D. A., Sakthikumar, S., Anderson, M. Z., Berlin, A., Gujja, S., Zeng, Q., Zisson, E., Wang, J. M., Greenberg, J. M., Berman, J., Bennett, R. J., and Cuomo, C. A. (2015). Genetic and phenotypic intra-species variation in Candida albicans. *Genome Research*, 25(3):413–425.

Hittinger, C. T., Rokas, A., Bai, F.-Y., Boekhout, T., Gonçalves, P., Jeffries, T. W., Kominek, J., Lachance, M.-A., Libkind, D., Rosa, C. A., Sampaio, J. P., and Kurtzman, C. P. (2015). Genomics and the making of yeast biodiversity. *Current opinion in genetics & development*, 35:100–9.

Ho, H.-l. and Haynes, K. (2015). Candida glabrata: new tools and technologies-expanding the toolkit. *FEMS yeast research*, 15(6):fov066.

Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J. C., Koch, R., Rauch, G.-J., White, S., Chow, W., Kilian, B., Quintais, L. T., Guerra-Assunção, J. A., Zhou, Y., Gu, Y., Yen, J., Vogel, J.-H., Eyre, T., Redmond, S., Banerjee, R., Chi, J., Fu, B., Langley, E., Maguire, S. F., Laird, G. K., Lloyd, D., Kenyon, E., Donaldson, S., Sehra, H., Almeida-King, J., Loveland, J., Trevanion, S., Jones, M., Quail, M., Willey, D., Hunt, A., Burton, J., Sims, S., McLay, K., Plumb, B., Davis, J., Clee, C., Oliver, K., Clark, R., Riddle, C., Eliott, D., Threadgold, G., Harden, G., Ware, D., Mortimer, B., Kerry, G., Heath, P., Phillimore, B., Tracey, A., Corby, N., Dunn, M., Johnson, C., Wood, J., Clark, S., Pelan, S., Griffiths, G., Smith, M., Glithero, R., Howden, P., Barker, N., Stevens, C., Harley, J., Holt, K., Panagiotidis, G., Lovell, J., Beasley, H., Henderson, C., Gordon, D., Auger, K., Wright, D., Collins, J., Raisen, C., Dyer, L., Leung, K., Robertson, L., Ambridge, K., Leongamornlert, D., McGuire, S., Gilderthorp, R., Griffiths, C., Manthravadi, D., Nichol, S., Barker, G., Whitehead, S., Kay, M., Brown, J., Murnane, C., Gray, E., Humphries, M., Sycamore, N., Barker, D., Saunders, D., Wallis, J., Babbage, A., Hammond, S., Mashreghi-Mohammadi, M., Barr, L., Martin, S., Wray, P., Ellington, A., Matthews, N., Ellwood, M., Woodmansey, R., Clark, G., Cooper, J., Tromans, A., Grafham, D., Skuce, C., Pandian, R., Andrews, R., Harrison, E., Kimberley, A., Garnett, J., Fosker, N., Hall, R., Garner, P., Kelly, D., Bird, C., Palmer, S., Gehring, I., Berger, A., Dooley, C. M., Ersan-Ürün, Z., Eser, C., Geiger, H., Geisler, M., Karotki, L., Kirn, A., Konantz, J., Konantz, M., Oberländer, M., Rudolph-Geiger, S., Teucke, M., Osoegawa, K., Zhu, B., Rapp, A., Widaa, S., Langford, C., Yang, F., Carter, N. P., Harrow, J., Ning, Z., Herrero, J., Searle, S. M. J., Enright, A., Geisler, R., Plasterk, R. H. A., Lee, C., Westerfield, M., de Jong, P. J., Zon, L. I., Postlethwait, J. H., Nüsslein-Volhard, C., Hubbard, T. J. P., Crollius, H. R., Rogers, J., Stemple, D. L., de Jong, P. J., Zon, L. I., Postlethwait, J. H., Nüsslein-Volhard, C., Hubbard, T. J. P., Roest Crollius, H., Rogers, J., and Stemple, D. L. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, 496(7446):498–503.

Hubisz, M. J., Falush, D., Stephens, M., and Pritchard, J. K. (2009). Inferring

weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9(5):1322–1332.

Huerta-Cepas, J., Capella-Gutiérrez, S., Pryszcz, L. P., Marcet-Houben, M., and Gabaldón, T. (2014). PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome. *Nucleic Acids Research*, 42(D1):D897–D902.

Hull, C. M. and Johnson, A. D. (1999). Identification of a mating type-like locus in the asexual pathogenic yeast Candida albicans. *Science (New York, N.Y.)*, 285(5431):1271–5.

Hunkapiller, T., Kaiser, R. J., Koop, B. F., and Hood, L. (1991). Large-scale and automated DNA sequence determination. *Science (New York, N.Y.)*, 254(5028):59–67.

Huson, D. H. and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular biology and evolution*, 23(2):254–67.

Iraqui, I., Garcia-Sanchez, S., Aubert, S., Dromer, F., Ghigo, J.-M., D'Enfert, C., and Janbon, G. (2004). The Yak1p kinase controls expression of adhesins and biofilm formation in Candida glabrata in a Sir4p-dependent pathway. *Molecular Microbiology*, 55(4):1259–1271.

James, T. Y., Kauff, F., Schoch, C. L., Matheny, P. B., Hofstetter, V., Cox, C. J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., Lumbsch, H. T., Rauhut, A., Reeb, V., Arnold, A. E., Amtoft, A., Stajich, J. E., Hosaka, K., Sung, G.-H., Johnson, D., O'Rourke, B., Crockett, M., Binder, M., Curtis, J. M., Slot, J. C., Wang, Z., Wilson, A. W., Schüßler, A., Longcore, J. E., O'Donnell, K., Mozley-Standridge, S., Porter, D., Letcher, P. M., Powell, M. J., Taylor, J. W., White, M. M., Griffith, G. W., Davies, D. R., Humber, R. A., Morton, J. B., Sugiyama, J., Rossman, A. Y., Rogers, J. D., Pfister, D. H., Hewitt, D., Hansen, K., Hambleton, S., Shoemaker, R. A., Kohlmeyer, J., Volkmann-Kohlmeyer, B., Spotts, R. A., Serdani, M., Crous, P. W., Hughes, K. W., Matsuura, K., Langer, E., Langer, G., Untereiner, W. A., Lücking, R., Büdel, B., Geiser, D. M., Aptroot, A., Diederich, P., Schmitt, I., Schultz, M., Yahr, R., Hibbett, D. S., Lutzoni, F., McLaughlin, D. J., Spatafora, J. W., and Vilgalys, R. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*, 443(7113):818–822.

Jeffroy, O., Brinkmann, H., Delsuc, F., and Philippe, H. (2006). Phylogenomics: the beginning of incongruence? *Trends in Genetics*, 22(4):225–231.

Jones, T., Federspiel, N. A., Chibana, H., Dungan, J., Kalman, S., Magee, B. B., Newport, G., Thorstenson, Y. R., Agabian, N., Magee, P. T., Davis, R. W., and Scherer, S. (2004). The diploid genome sequence of Candida albicans.

Proceedings of the National Academy of Sciences of the United States of America, 101(19):7329–34.

Karpov, S. A., Mamkaeva, M. A., Aleoshin, V. V., Nassonova, E., Lilje, O., and Gleason, F. H. (2014). Morphology, phylogeny, and ecology of the aphelids (Aphelidea, Opisthokonta) and proposal for the new superphylum Opisthosporidia. *Frontiers in Microbiology*, 5:112.

Katiyar, S. K., Alastruey-Izquierdo, A., Healey, K. R., Johnson, M. E., Perlin, D. S., and Edlind, T. D. (2012). Fks1 and Fks2 Are Functionally Redundant but Differentially Regulated in Candida glabrata: Implications for Echinocandin Resistance. *Antimicrobial Agents and Chemotherapy*, 56(12):6304–6309.

Katoh, K., Misawa, K., Kuma, K.-i., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, 30(14):3059–66.

Kaur, R., Domergue, R., Zupancic, M. L., and Cormack, B. P. (2005). A yeast by any other name: Candida glabrata and its interaction with the host. *Current Opinion in Microbiology*, 8(4):378–384.

Kaur, R., Ma, B., and Cormack, B. P. (2007). A family of glycosylphosphatidylinositol-linked aspartyl proteases is required for virulence of Candida glabrata. *Proceedings of the National Academy of Sciences of the United States of America*, 104(18):7628–33.

Kellis, M., Birren, B. W., and Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature*, 428(6983):617–624.

Koonin, E. V. (2005). Orthologs, Paralogs, and Evolutionary Genomics. *Annual Review of Genetics*, 39(1):309–338.

Kurtzman, C. P. (2003). Phylogenetic circumscription of Saccharomyces, Kluyveromyces and other members of the Saccharomycetaceae, and the proposal of the new genera Lachancea, Nakaseomyces, Naumovia, Vanderwaltozyma and Zygotorulaspora. *FEMS yeast research*, 4(3):233–45.

Kurtzman, C. P., Fell, J. W., and Boekhout, T. (2011). *The yeasts : a taxonomic study*. Elsevier.

Kurtzman, C. P. and Robnett, C. J. (1998). Identification and phylogeny of ascomycetous yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences. *Antonie van Leeuwenhoek*, 73(4):331–71.

Lafontaine, I., Fischer, G., Talla, E., and Dujon, B. (2004). Gene relics in the genome of the yeast Saccharomyces cerevisiae. *Gene*, 335:1–17.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics (Oxford, England)*, 27(21):2987–93.

Li, H. and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 26(5):589–95.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16):2078–9.

Li, N. and Stephens, M. (2003). Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*, 165(4):2213–33.

Li, Y.-F. and Bao, W.-G. (2007). Why do some yeast species require niacin for growth? Different modes of NAD synthesis. *FEMS Yeast Research*, 7(5):657–664.

Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., Davey, R. P., Roberts, I. N., Burt, A., Koufopanou, V., Tsai, I. J., Bergman, C. M., Bensasson, D., O'Kelly, M. J. T., van Oudenaarden, A., Barton, D. B. H., Bailes, E., Nguyen, A. N., Jones, M., Quail, M. A., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R., and Louis, E. J. (2009). Population genomics of domestic and wild yeasts. *Nature*, 458(7236):337–41.

Lockhart, S. R., Iqbal, N., Cleveland, A. A., Farley, M. M., Harrison, L. H., Bolden, C. B., Baughman, W., Stein, B., Hollick, R., Park, B. J., and Chiller, T. (2012). Species Identification and Antifungal Susceptibility Testing of Candida Bloodstream Isolates from Population-Based Surveillance Studies in Two U.S. Cities from 2008 to 2011. *Journal of Clinical Microbiology*, 50(11):3435–3442.

Lockhart, S. R., Messer, S. A., Gherna, M., Bishop, J. A., Merz, W. G., Pfaller, M. A., and Diekema, D. J. (2009). Identification of Candida nivariensis and Candida bracarensis in a Large Global Collection of Candida glabrata Isolates: Comparison to the Literature. *Journal of Clinical Microbiology*, 47(4):1216–1217.

Luikart, G., England, P. R., Tallmon, D., Jordan, S., and Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, 4(12):981–994.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D. W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.-W., and Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 1(1):18.

MacCallum, D. M., Findon, H., Kenny, C. C., Butler, G., Haynes, K., and Odds, F. C. (2006). Different Consequences of ACE2 and SWI5 Gene Disruptions for Virulence of Pathogenic and Nonpathogenic Yeasts. *Infection and Immunity*, 74(9):5244–5248.

Maillet, L., Boscheron, C., Gotta, M., Marcand, S., Gilson, E., and Gasser, S. M. (1996). Evidence for silencing compartments within the yeast nucleus: a role for telomere proximity and Sir protein concentration in silencer-mediated repression. *Genes & development*, 10(14):1796–811.

Malani, A. N., Psarros, G., Malani, P. N., and Kauffman, C. A. (2011). Is age a risk factor for Candida glabrata colonisation? *Mycoses*, 54(6):531–537.

Marcet-Houben, M. and Gabaldón, T. (2010). Acquisition of prokaryotic genes by fungal genomes. *Trends in Genetics*, 26(1):5–8.

Marcet-Houben, M. and Gabaldón, T. (2015). Beyond the whole-genome duplication: Phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLoS Biology*, 13(8):1–26.

Marichal, P., Odds, F. C., Ramaekers, F. C. S., Bellens, D., Vanden Bossche, H., Luyten, W., Borgers, M., Willemsens, S., Koymans, L., and Verhasselt, P. (1999). Contribution of mutations in the cytochrome P450 14α-demethylase (Erg11p, Cyp51p) to azole resistance in Candida albicans. *Microbiology*, 145(10):2701–2713.

Martin, D. and Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics (Oxford, England)*, 16(6):562–3.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2009). RDP4: Detection and analysis of recombination patterns in virus genomes.

Mason, K. L., Erb Downward, J. R., Mason, K. D., Falkowski, N. R., Eaton, K. A., Kao, J. Y., Young, V. B., and Huffnagle, G. B. (2012). Candida albicans and Bacterial Microbiota Interactions in the Cecum during Recolonization following Broad-Spectrum Antibiotic Therapy. *Infection and Immunity*, 80(10):3371–3380.

Mayer, F. L., Wilson, D., and Hube, B. (2013). Candida albicans pathogenicity mechanisms. *Virulence*, 4(2):119–128.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9):1297–303.

McVean, G. A. T., Myers, S. R., Hunt, S., Deloukas, P., Bentley, D. R., and Donnelly, P. (2004). The fine-scale structure of recombination rate variation in the human genome. *Science (New York, N.Y.)*, 304(5670):581–4.

Moreira, A., Silva, S., Botelho, C., Sampaio, P., Pais, C., and Henriques, M. (2015). Candida bracarensis: Evaluation of Virulence Factors and its Tolerance to Amphotericin B and Fluconazole. *Mycopathologia*, 180(5-6):305–315.

Mostowy, R., Croucher, N. J., Andam, C. P., Corander, J., Hanage, W. P., and Marttinen, P. (2017). Efficient Inference of Recent and Ancestral Recombination within Bacterial Populations. *Molecular biology and evolution*, 34(5):1167–1182.

Mühlhausen, S. and Kollmar, M. (2014). Molecular phylogeny of sequenced Saccharomycetes reveals polyphyly of the alternative yeast codon usage. *Genome biology and evolution*, 6(12):3222–37.

Muller, H., Hennequin, C., Gallaud, J., Dujon, B., and Fairhead, C. (2008). The asexual yeast Candida glabrata maintains distinct a and α haploid mating types. *Eukaryotic Cell*, 7(5):848–858.

Muller, H., Thierry, A., Coppée, J.-Y., Gouyette, C., Hennequin, C., Sismeiro, O., Talla, E., Dujon, B., and Fairhead, C. (2009). Genomic polymorphism in the population of Candida glabrata: gene copy-number variation and chromosomal translocations. *Fungal genetics and biology : FG & B*, 46(3):264–76.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A. A., Korobeynikov, A., Lapidus, A., Prjibelski, A. D., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., Clingenpeel, S. R., Woyke, T., Mclean, J. S., Lasken, R., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (2013). Assembling Single-Cell Genomes and Mini-Metagenomes From Chimeric MDA Products. *Journal of Computational Biology*, 20(10):714–737.

Oliver, B. G., Song, J. L., Choiniere, J. H., and White, T. C. (2007). cis-Acting Elements within the Candida albicans ERG11 Promoter Mediate the Azole Response through Transcription Factor Upc2p. *Eukaryotic Cell*, 6(12):2231–2239.

Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS genetics*, 2(12):e190.

Perlin, D. S., Shor, E., and Zhao, Y. (2015). Update on Antifungal Drug Resistance. *Current Clinical Microbiology Reports*, 2(2):84–95.

Petersen, J. H. (2013). *The kingdom of fungi*.

Pfaller, M. A. and Diekema, D. J. (2007). Epidemiology of invasive candidiasis: a persistent public health problem. *Clinical microbiology reviews*, 20(1):133–63.

Pfaller, M. A., Diekema, D. J., Gibbs, D. L., Newell, V. A., Barton, R., Bijie, H., Bille, J., Chang, S.-C., da Luz Martins, M., Duse, A., Dzierzanowska, D., Ellis, D., Finquelievich, J., Gould, I., Gur, D., Hoosen, A., Lee, K., Mallatova, N., Mallie, M., Peng, N. K., Petrikos, G., Santiago, A., Trupl, J., VanDen Abeele, A. M., Wadula, J., Zaidi, M., and Global Antifungal Surveillance Group (2010). Geographic variation in the frequency of isolation and fluconazole and voriconazole susceptibilities of Candida glabrata: an assessment from the ARTEMIS DISK Global Antifungal Surveillance Program. *Diagnostic Microbiology and Infectious Disease*, 67(2):162–171.

Pfaller, M. A., Messer, S. A., Boyken, L., Tendolkar, S., Hollis, R. J., and Diekema, D. J. (2003). Variation in susceptibility of bloodstream isolates of Candida glabrata to fluconazole according to patient age and geographic location. *Journal of clinical microbiology*, 41(5):2176–9.

Pfaller, M. A., Messer, S. A., Hollis, R. J., Boyken, L., Tendolkar, S., Kroeger, J., and Diekema, D. J. (2009). Variation in Susceptibility of Bloodstream Isolates of Candida glabrata to Fluconazole According to Patient Age and Geographic Location in the United States in 2001 to 2007. *Journal of Clinical Microbiology*, 47(10):3185–3190.

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., and Lercher, M. J. (2014). PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Molecular biology and evolution*, 31(7):1929–36.

Philippe, H., Delsuc, F., Brinkmann, H., and Lartillot, N. (2005). Phylogenomics. *Annual Review of Ecology, Evolution, and Systematics*, 36(1):541–562.

Plaut, A. (1950). Human infection with Cryptococcus glabratus; report of case involving uterus and fallopian tube. *American journal of clinical pathology*, 20(4):377–80.

Pool, J. E., Hellmann, I., Jensen, J. D., and Nielsen, R. (2010). Population genetic inference from genomic sequence variation. *Genome research*, 20(3):291–300.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, 38(8):904–9.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959.

Pryszcz, L. P., Huerta-Cepas, J., and Gabaldón, T. (2011). MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score. *Nucleic Acids Research*, 39(5):e32–e32.

Pundir, S., Magrane, M., Martin, M. J., O'Donovan, C., and UniProt Consortium (2015). Searching and Navigating UniProt Databases. In *Current Protocols in Bioinformatics*, volume 50, pages 1.27.1–1.27.10. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Rai, M. N., Balusu, S., Gorityala, N., Dandu, L., and Kaur, R. (2012). Functional Genomic Analysis of Candida glabrata-Macrophage Interaction: Role of Chromatin Remodeling in Virulence. *PLoS Pathogens*, 8(8):e1002863.

Redecker, D. and Raab, P. (2006). Phylogeny of the glomeromycota (arbuscular mycorrhizal fungi): recent developments and new gene markers. *Mycologia*, 98(6):885–95.

Richard, G. F., Kerrest, A., Lafontaine, I., and Dujon, B. (2005). Comparative genomics of hemiascomycete yeasts: Genes involved in DNA replication, repair, and recombination. *Molecular Biology and Evolution*, 22(4):1011–1023.

Richards, T. A., Leonard, G., and Wideman, J. G. (2017). What Defines the Kingdom Fungi? *Microbiology Spectrum*, 5(3).

Richardson, M. and Lass-Flörl, C. (2008). Changing epidemiology of systemic fungal infections. *Clinical Microbiology and Infection*, 14:5–24.

Rodrigues, C. F., Silva, S., and Henriques, M. (2014). Candida glabrata: a review of its features and resistance. *European Journal of Clinical Microbiology & Infectious Diseases*, 33(5):673–688.

Roetzer, A., Gabaldón, T., and Schüller, C. (2011). From Saccharomyces cerevisiaeto Candida glabratain a few easy steps: important adaptations for an opportunistic pathogen. *FEMS Microbiology Letters*, 314(1):1–9.

Roetzer, A., Gratz, N., Kovarik, P., and SchÃller, C. (2010). Autophagy supports Candida glabrata survival during phagocytosis. *Cellular Microbiology*, 12(2):199–216.

Salvadó, Z., Arroyo-López, F. N., Guillamón, J. M., Salazar, G., Querol, A., and Barrio, E. (2011). Temperature adaptation markedly determines evolution within the genus Saccharomyces. *Applied and environmental microbiology*, 77(7):2292–302.

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., Hutchison, C. A., Slocombe, P. M., and Smith, M. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*, 265(5596):687–95.

Santos, M. A. S., Gomes, A. C., Santos, M. C., Carreto, L. C., and Moura, G. R. (2011). The genetic code of the fungal CTG clade. *Comptes Rendus - Biologies*, 334(8-9):607–611.

Scazzocchio, C. (2014). Fungal biology in the post-genomic era. *Fungal Biology and Biotechnology*, 1(1):7.

Schwarzmüller, T., Ma, B., Hiller, E., Istel, F., Tscherner, M., Brunke, S., Ames, L., Firon, A., Green, B., Cabral, V., Marcet-Houben, M., Jacobsen, I. D., Quintin, J., Seider, K., Frohner, I., Glaser, W., Jungwirth, H., Bachellier-Bassi, S., Chauvel, M., Zeidler, U., Ferrandon, D., Gabaldón, T., Hube, B., D'Enfert, C., Rupp, S., Cormack, B., Haynes, K., and Kuchler, K. (2014). Systematic phenotyping of a large-scale Candida glabrata deletion collection reveals novel antifungal tolerance genes. *PLoS pathogens*, 10(6):e1004211.

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

Selmecki, A., Forche, A., and Berman, J. (2010). Genomic Plasticity of the Human Fungal Pathogen Candida albicans. *Eukaryotic Cell*, 9(7):991–1008.

Selmecki, A., Gerami-Nejad, M., Paulson, C., Forche, A., and Berman, J. (2008). An isochromosome confers drug resistance in vivo by amplification of two genes, ERG11 and TAC1. *Molecular Microbiology*, 68(3):624–641.

Selmecki, A. M., Dulmage, K., Cowen, L. E., Anderson, J. B., and Berman, J. (2009). Acquisition of Aneuploidy Provides Increased Fitness during the Evolution of Antifungal Drug Resistance. *PLoS Genetics*, 5(10):e1000705.

Shendure, J. and Aiden, E. L. (2012). The expanding scope of DNA sequencing. *Nature Biotechnology*, 30(11):1084–1094.

Shendure, J. and Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, 26(10):1135–1145.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic acids research*, 29(1):308–11.

Slatkin, M. (2008). Linkage disequilibrium understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, 9(6):477–485.

Smith, S. E. and Read, D. J. D. J. (2008). *Mycorrhizal symbiosis*. Academic Press.

Spatafora, J. W., Chang, Y., Benny, G. L., Lazarus, K., Smith, M. E., Berbee, M. L., Bonito, G., Corradi, N., Grigoriev, I., Gryganskyi, A., James, T. Y., O'Donnell, K., Roberson, R. W., Taylor, T. N., Uehling, J., Vilgalys, R., White, M. M., and Stajich, J. E. (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia*, 108(5):1028–1046.

Sprouffske, K. and Wagner, A. (2016). Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC bioinformatics*, 17(1):172.

Srikantha, T., Lachke, S. A., and Soll, D. R. (2003). Three mating type-like loci in Candida glabrata. *Eukaryotic cell*, 2(2):328–40.

Stajich, J. E., Berbee, M. L., Blackwell, M., Hibbett, D. S., James, T. Y., Spatafora, J. W., and W., T. J. (2009). The Fungi. *Current Biology*, 19(18):R840–R845.

Stamatakis, A., Ludwig, T., and Meier, H. (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics (Oxford, England)*, 21(4):456–463.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic acids research*, 34(Web Server issue):W435–9.

Sugita, T. and Nakase, T. (1999). Non-universal Usage of the Leucine CUG Codon and the Molecular Phylogeny of the Genus Candida. *Systematic and Applied Microbiology*, 22(1):79–86.

Swerdlow, H., Wu, S. L., Harke, H., and Dovichi, N. J. (1990). Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. *Journal of chromatography*, 516(1):61–7.

Tavanti, A., Davidson, A. D., Fordyce, M. J., Gow, N. A. R., Maiden, M. C. J., and Odds, F. C. (2005). Population structure and properties of Candida albicans, as determined by multilocus sequence typing. *Journal of clinical microbiology*, 43(11):5601–13.

Tenenhaus, M. and Young, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika*, 50(1):91–119.

Tsai, H.-F., Krol, A. A., Sarti, K. E., and Bennett, J. E. (2006). Candida glabrata PDR1, a Transcriptional Regulator of a Pleiotropic Drug Resistance Network, Mediates Azole Resistance in Clinical Isolates and Petite Mutants. *Antimicrobial Agents and Chemotherapy*, 50(4):1384–1392.

Vale-Silva, L., Beaudoing, E., Tran, V. D. T., and Sanglard, D. (2017). Comparative Genomics of Two Sequential Candida glabrata Clinical Isolates. *G3: Genes—Genomes—Genetics*, page g3.117.042887.

Vale-Silva, L. A. and Sanglard, D. (2015). Tipping the balance both ways: drug resistance and virulence in Candida glabrata. *FEMS Yeast Research*, 15(4):fov025.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., and DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]*, 43(1110):11.10.1–33.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9):418–426.

Watkinson, S. C., Boddy, L., and Money, N. P. (2016). *The Fungi (Third Edition)*. Elsevier.

Whitaker, J. W., McConkey, G. A., and Westhead, D. R. (2009). The transferome of metabolic genes explored: analysis of the horizontal transfer of enzyme encoding genes in unicellular eukaryotes. *Genome biology*, 10(4):R36.

White, T. C., Holleman, S., Dy, F., Mirels, L. F., and Stevens, D. A. (2002). Resistance mechanisms in clinical isolates of Candida albicans. *Antimicrobial agents and chemotherapy*, 46(6):1704–13.

White, T. C., Marr, K. A., and Bowden, R. A. (1998). Clinical, cellular, and molecular factors that contribute to antifungal drug resistance. *Clinical microbiology reviews*, 11(2):382–402.

Wickham, H. (2009). *ggplot2*. Springer New York, New York.

Wilson, L. S., Reyes, C. M., Stolpman, M., Speckman, J., Allen, K., and Beney, J. (2002). The Direct Cost and Incidence of Systemic Fungal Infections. *Value in Health*, 5(1):26–34.

Wolfe, K. H. and Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, 387(6634):708–713.

Wong, S., Fares, M. A., Zimmermann, W., Butler, G., and Wolfe, K. H. (2003). Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast Candida glabrata. *Genome biology*, 4(2):R10.

Xu, N., Ye, C., Chen, X., Liu, J., Liu, L., and Chen, J. (2016). Genome Sequencing of the Pyruvate-producing Strain Candida glabrata CCTCC M202019 and Genomic Comparison with Strain CBS138. *Scientific reports*, 6(1):34893.

Yáñez-Carrillo, P., Robledo-Márquez, K. A., Ramírez-Zavaleta, C. Y., De Las Peñas, A., and Castaño, I. (2014). The mating type-like loci of Candida glabrata. *Revista iberoamericana de micologia*, 31(1):30–4.

Zuckerkandl, E. and Pauling, L. (1965). Molecules as documents of evolutionary history. *Journal of theoretical biology*, 8(2):357–66.