Review

Molecular Epidemiology, Phylogeny and Evolution of *Candida albicans*

Brenda A. McManus [a], David C. Coleman [a*]

[a] Microbiology Research Unit, Division of Oral Biosciences, Dublin Dental University
Hospital, Trinity College Dublin, University of Dublin, Dublin 2, Republic of Ireland

[*]Corresponding author: Mailing address: Microbiology Research Unit, Division of Oral Biosciences, Dublin Dental University Hospital, University of Dublin, Trinity College Dublin, Dublin 2, Republic of Ireland. Phone: +353 1 6127276. Fax: +353 1 6127295. E-mail : david.coleman@dental.tcd.ie.

ABSTRACT (300 words max)

A small number of *Candida* species form part of the normal microbial flora of mucosal surfaces in humans and may give rise to opportunistic infections when host defences are impaired. *Candida albicans* is by far the most prevalent commensal and pathogenic *Candida* species. Several different molecular typing approaches including multilocus sequence typing, multilocus microsatellite typing and DNA fingerprinting using *C. albicans*-specific repetitive sequence- containing DNA probes have yielded a wealth of information regarding the epidemiology and population structure of this species. Such studies revealed that the *C. albicans* population structure consists of multiple major and minor clades, some of which exhibit geographical or phenotypic enrichment and that *C. albicans* reproduction is predominantly clonal. Despite this, losses of heterozygosity by recombination, the existence of a parasexual cycle, toleration of a wide range of aneuploidies and the recent description of viable haploid strains have all demonstrated the extensive plasticity of the *C. albicans* genome. Recombination and gross chromosomal rearrangements are more common under stressful environmental conditions, and have played a significant role in the evolution of this opportunistic pathogen.

Surprisingly, *Candida dubliniensis*, the closest relative of *C. albicans* exhibits more karyotype variability than *C. albicans*, but is significantly less adaptable to unfavourable environments. This disparity most likely reflects the evolutionary processes that occurred during or soon after the divergence of both species from their common ancestor. Whilst *C. dubliniensis* underwent significant gene loss and pseudogenisation, *C. albicans* expanded gene families considered to be important in virulence.

It is likely that technological developments in whole genome sequencing and data analysis in coming years will facilitate its routine use for population structure, epidemiological investigations, and phylogenetic analyses of *Candida* species. These are likely to reveal more minor *C. albicans* clades and to enhance our understanding of the population biology of this versatile organism.

1. Introduction

Candida species are typically harmless eukaryotic commensal yeasts that are members of the phylum Ascomycota and can be recovered from environmental, human and other mammalian sources. In mammals, *Candida* species most commonly reside as part of the normal commensal microbial flora on mucosal surfaces of the gastrointestinal and genitourinary tracts (Kumamoto, 2011) in healthy individuals and cause infection only when host immunity becomes compromised. The prevalence of *Candida* is higher in pregnant, diabetic, elderly or immunocompromised individuals, or in those who wear dentures or are receiving antibiotic or corticosteroid treatment, and these are predisposing factors for *Candida* infection (Lockhart et al., 2003; Odds, 1988).

Only a relatively small number of *Candida* species are of clinical importance in humans including *Candida albicans*, *Candida glabrata*, *Candida tropicalis*, *Candida parapsilosis* and *Candida dubliniensis*. *Candida albicans* is the most prevalent and the most pathogenic of the *Candida* species, and is responsible for the majority of oral and systemic candidiasis cases (Moran et al., 2012; Thompson et al., 2010; Zomorodian et al., 2011) as well as community-onset and nosocomial candidaemias (Pfaller et al., 2010). *Candida albicans* is a dimorphic species that can grow as yeast or filamentous forms, and is one of the only two *Candida* species capable of forming true hyphae, the other species being *C. dubliniensis*, the closest relative of *C. albicans*. Hyphae are considered to play important roles in processes such as adhesion and tissue invasion. Comparison of both species in both mucosal and systemic infection models have demonstrated that in spite of the ability of both species to produce hyphae, *C. albicans* is a significantly more successful pathogen (Asmundsdottir et al., 2009; Stokes et al., 2007).

The *C. albicans* genome consists of eight pairs of chromosomal homologs, ranging in size from 0.95 to 3.3 Mb in size and comprising 16 Mb in total (Chibana et al., 2000). The species is predominantly diploid, however it exhibits a high degree of genome plasticity and exhibits frequent losses of heterozygosity as well as gross chromosomal rearrangements that may result in aneuploidy. Though reproduction is predominantly clonal, the species can also utilise a parasexual cycle involving the formation of tetraploid progeny from the mating of diploid parents, the former of which subsequently revert to diploidy by concerted chromosome loss (Bennett and Johnson, 2003). The parasexual cycle occurs rarely in nature, possibly only under stressful conditions and is further described in section 4.1 below.

The purpose of this short article is to succinctly review the molecular epidemiology and phylogeny of *C. albicans* by examining its population structure, the plasticity of its genome, evolutionary pathway and ancestry.

2. Molecular typing of *C. albicans*

Molecular typing systems have proved very useful in the epidemiological and population structure analyses of microbial pathogens, facilitating the understanding of the dynamics of infectious organisms in human populations, the complex relationships between commensal and infectious organisms, the origins of infection, the emergence of drug resistance in populations and the genetic relatedness of isolates of the same species. Informative molecular typing systems tend to share specific features that enable them to be used effectively for isolate discrimination and for determining a measure of isolate relatedness. Ideally, such typing systems have to be (i) effective at discriminating between closely related but non-identical isolates of the same species; (ii) be able to recognise the same strain among collections of different isolates and generate reproducible data; (iii) be unaffected by high frequency genome reorganisation and evolutionary pressure so that genetic differences are relatively stable over time and mutate with a medium frequency, thus reflecting evolutionary change only; (iv) be able to determine the genetic distance between isolates that are more closely related and those that are less so and (v) be amenable to computer-based analysis to facilitate data normalisation, analysis and storage. Several comprehensive reviews of molecular methods for typing of *C. albicans* have been published (Gil-Lamaignere et al., 2003; Saghrouni et al., 2013; Soll, 2000) and therefore, only a brief overview is provided here.

2.1 Earlier molecular-based typing methods

Prior to the advent of molecular typing techniques, studies of the epidemiology of *C. albicans* relied on phenotypic methods such as morphotyping, serotyping, biotyping and antimicrobial agent susceptibility testing. These approaches were limited by poor inter-laboratory reproducibility and/or poor discriminatory ability. Multi-locus enzyme electrophoresis (MLEE) was developed as a more reproducible and discriminatory phenotypic typing technique based on the differential electrophoretic mobility of approximately ten enzymatic proteins (Caugant and Sandven, 1993). The mobility of these proteins varies according to altered molecular size and net charge resulting from amino acid substitutions, and therefore, this technique is only able to detect major evolutionary changes amongst isolates. MLEE has largely been replaced with more discriminatory typing methods that target DNA more directly such as restriction fragment length polymorphism (RFLP) analysis, electrophoretic karyotyping (EK), and random amplified polymorphic DNA (RAPD) analysis (Soll, 2000). These methods can detect alterations in the DNA sequences of restriction endonuclease cleavage sites that may be caused by single nucleotide polymorphisms (SNPs),

insertions/deletions, translocations, recombination events or transposable element activity. Similarly to RAPD, PCR fingerprinting using one or more arbitrary primers has been used to detect intra-species variability of *C. albicans* (Meyer et al., 1993; Bartie et al., 2001). However, all these methods are somewhat limited due to poor inter-laboratory reproducibility.

Prior to the application of direct DNA sequencing methods for typing, the most widely used DNA-based typing method for population structure analysis of *C. albicans* involved Southern hybridisation analysis of restriction endonuclease-digested chromosomal DNA using complex species-specific DNA fingerprinting probes such as 27A and Ca3, consisting of cloned chromosomal DNA fragments homologous to repetitive DNA sequences dispersed throughout the *C. albicans* genome (Pujol et al., 1997; Scherer and Stevens, 1988). The application of the highly discriminatory Ca3 DNA fingerprinting probe to *C. albicans* isolates supported the grouping of three distinct but geographically widespread clades (clades I-III) that are also identified by MLEE and RAPD (Robles et al., 2004). The Ca3 fingerprinting studies also revealed two additional clades that were geographically enriched with isolates from South Africa and Europe, termed the SA and E clades, respectively (Blignaut et al., 2002; Boerlin et al., 1995; Pujol et al., 2002; Scherer and Stevens, 1988; Soll and Pujol, 2003). The discriminatory power of a typing method is determined according to the average probability that the method will assign a different type to two randomly sampled and unrelated strains. Whilst DNA fingerprinting using complex DNA probes has a very high discriminatory power of 0.993 (Robles et al., 2004), the method is laborious, technically demanding and inter-laboratory comparisons are difficult, making methods that rely on the direct analysis of DNA sequences more attractive and more viable alternatives.

2.2 Typing methods based on DNA sequence comparisons

The more widespread availability of high-throughput and affordable DNA sequencing technology in the last two decades and the determination of whole genome DNA sequences from a range of *C. albicans* strains revolutionised molecular typing of this organism, enabling the development of typing systems based on the direct comparison of DNA sequences. Over the last ten years DNA-sequence based approaches have been applied to epidemiological studies of *C. albicans* with great success, and have provided a wealth of information regarding the population structure of the species. A variety of molecular typing systems based on direct DNA sequence comparison have been developed and applied to *C. albicans*.

Multilocus microsatellite typing (MLMT) targets codominantly inherited stretches of tandemly repeated sequences that exhibit considerable hypervariability in quantities of repeats amongst isolates (Sampaio et al., 2005). Repeated areas are amplified by polymerase chain

reaction (PCR) using primers that flank a specific microsatellite region in the genome and allele sizes are determined by automated capillary electrophoresis. Due to its high discriminatory capacity of 0.987 (Chavez-Galarza et al., 2010), this technique has been used for strain typing (Dalle et al., 2000), population structure analyses (Chavez-Galarza et al., 2010; Fundyga et al., 2002) and epidemiological studies (Fundyga et al., 2002; Sampaio et al., 2003) of *C. albicans*. Whilst this method is highly discriminatory and reproducible, the attractiveness of the technique is diminished by the lack of a standardised *C. albicans* MLMT typing scheme and associated publicly accessible and curated online database.

Multilocus sequence typing (MLST) of *C. albicans* isolates involves the PCR amplification and DNA sequence analysis of 300-400 bp regions from seven housekeeping genes that are under stabilising selection pressure (Bougnoux et al., 2002; Bougnoux et al., 2003; Tavanti et al., 2003). For each locus, sequence variations caused by SNPs are identified as separate alleles. Each unique allele is assigned a corresponding integer, and each unique combined set of integers per isolate (an allelic profile) is assigned another corresponding integer, defined as a sequence type (ST). The diploid nature of *C. albicans* increases the level of sequence variation due to the presence of heterozygous nucleotide sites, which provide additional genotypes. In *C. albicans*, the ST is also referred to as a diploid sequence type (DST). As it is based on direct DNA sequence comparisons, MLST is highly reproducible and has a discriminatory power of 0.999 that is comparable to the discriminatory power (0.993) of Ca3-based fingerprinting (Odds & Jacobsen, 2008; Robles et al., 2004). Most importantly, MLST data are directly comparable amongst different research groups around the world (Fig. 1), enabling collaborative studies *via* the publicly accessible online curated *C. albicans* MLST database (http://calbicans.mlst.net/).

Computer-based analysis of DNA sequence or allelic profile data can be used to generate phylogenetic trees based on unweighted pair group method with arithmetic averages (UPGMA), maximum parsimony and neighbour-joining methods to display the genetic relatedness of the isolates being investigated. An alternative algorithm based upon related sequence types (BURST) was originally designed by Feil and Enright (2004) for inferring genetic relationships and founding genotypes for clonal complexes (CCs) amongst bacterial species and adapted for use with MLST data (Feil et al., 2004). The BURST algorithm has been applied to allelic profile data from the MLST analysis of *C. albicans* and used to infer which DSTs may have been founding genotypes for each CC (Fig. 2). Whilst the BURST method does not provide a better classification than the other clustering methods, the different algorithms define clades with a relatively good correlation (Odds and Jacobsen 2008) and many *C. albicans*-based MLST studies use the BURST algorithm as a secondary clustering

method (Ge et al., 2012; Gong et al., 2012; Odds et al., 2007; Takakura et al., 2008; Tavanti et al., 2005).

In the last decade, MLST has become the most popular method for the molecular typing, (Bougnoux et al., 2002; Odds et al., 2007), population structure analysis (Bougnoux et al., 2008b; Odds et al., 2007; Tavanti et al., 2005) and epidemiological analysis (Bougnoux et al., 2006; Bougnoux et al., 2008a; Chen et al., 2006; Da Matta et al., 2010; Gammelsrud et al., 2012; McManus et al., 2012; McManus et al., 2011) of *C. albicans*. Interestingly, MLST can be used to carry out comparative population structure analysis on closely related species, and one such study of the close relatives *C. albicans* and *C. dubliniensis* revealed that the population structure of *C. albicans* is significantly more divergent than that of *C. dubliniensis*, correlating with the lower prevalence and pathogenic potential of *C. dubliniensis* (McManus et al., 2008; McManus et al., 2009).

3. Molecular Epidemiology and Population Structure of *C. albicans*

Several epidemiological studies using MLST and microsatellite typing have shown that *C. albicans* infections very often arise from an endogenous source and persistent *C. albicans* strains are maintained by hosts over prolonged periods of time, occasionally undergoing minor genetic variations known as microvariation (Bougnoux et al., 2006; Da Matta et al., 2010; Gammelsrud et al., 2012; Jacobsen et al., 2008b; McManus et al., 2011; Sampaio et al., 2005; Stephan et al., 2002). Strain replacement has also been observed, as has the transmission of *C. albicans* isolates between different individuals and microvariation of persistent isolates occurring in the same individual between recurrent infections (Bougnoux et al., 2006; Cliff et al., 2008; McManus et al., 2011; Odds et al., 2006; Shin et al., 2011). Co-dominant markers based on SNPs in PCR amplicons have been used to demonstrate that the population of *C. albicans* is predominantly clonal, but the data generated also suggests recombination events may occur, albeit rarely (Gräser et al., 1996; Forche et al., 1999).

The population structure of *C. albicans* appears to be quite robust, as Ca3 fingerprinting, MLMT, and MLST all show good correlation in the definition of major clades. Previously identified Ca3 clades I, II, III and SA correspond precisely with MLST clades 1– 4 respectively, and these clades also emerged as distinct groups according to MLMT using the CA1 and CEF3 microsatellite regions as markers (Chavez-Galarza et al., 2010). These clades also emerge upon analysis of isolates recovered exclusively from one geographical location, suggesting that these groups are deeply rooted and are genuinely genetically distinct from each other (Bougnoux et al., 2008b; Odds et al., 2007; Soll, 2000). Comprehensive analysis of a population of 1,391 *C. albicans* isolates by MLST identified 17 clades using an arbitrary *P*

distance (the proportion of all nucleotides examined that exhibit polymorphism) of 0.04 as a threshold to delineate clades (Odds et al., 2007). An additional clade enriched with isolates from Asia was more recently proposed by Shin et al. (2011). These more minor clades have not been described using MLMT to date, most likely due to the application of a different threshold (0.7) for the purposes of defining the five major clades, and due to the lesser number of isolates typed by MLMT.

Throughout these clades, isolates can be further distinguished according to ABC genotypes based on the absence, presence or heterozygous presence of a transposable intron in the internal transcribed spacer region of the 25S rRNA gene (McCullough et al., 1999). These genotypes are referred to as ABC genotypes A, B and C, respectively (Table 1). Odds et al., (2007) observed that individual MLST clades were enriched with particular ABC genotypes or with isolates recovered from specific geographical locations (Table 1), correlating with previous studies (Tavanti et al., 2005) based on MLST. Clade-specific enrichment of isolates recovered predominantly from South Africa and Europe were first revealed using Ca3-based DNA fingerprinting analyses (Blignaut et al., 2002; Pujol et al., 2003, however these earlier studies did not include ABC genotyping analyses.

MLST clade 1 appears to have a global distribution, clade 2 is enriched with isolates recovered from the United Kingdom, clade 4 is enriched with isolates from the Middle East and Africa, clade 11 is enriched with isolates from continental Europe, and isolates recovered from the Pacific Rim tend to cluster in clades 14 and 17 (Odds et al., 2007) (Table 1). More recently, MLST isolates recovered from bloodstream infections in South Korea suggest the emergence of a novel Asian clade (Shin et al., 2011) (Fig. 2).

Upon removal of potential geographical effects by only examining data from isolates recovered in Western Europe (*n*=559), Odds et al., (2007) were able to demonstrate significant clade distributions among bloodstream, commensal and superficial infection-causing isolates. This supported earlier findings by Tavanti et al. (2005) based on MLST analysis of 395 isolates. Both of these studies reported statistically significant differential distribution of blood, vaginal, and oropharyngeal isolates amongst different predominant MLST clades. Isolates exhibiting reduced 5-fluorocytosine (5FC) susceptibility significantly clustered in MLST clade 1 (Odds et al., 2007; Odds and Jacobsen, 2008; Tavanti et al., 2005), the most predominant clade, due to the clonal spread of a co-dominant R101C transition in the *FUR1* gene (Dodgson et al., 2004) that encodes a uracil phosphoribosyl transferase. A significant association of salt tolerance with isolates belonging to clade 1 has also been reported (MacCallum et al., 2009). The same study described a significantly lower level of acid phosphatase activity in isolates belonging to MLST clade 2, as well as clade-specific

associations of numbers of tandem repeat sequences in the hyphal regulator genes (*HYR1* and *HYR3*) and the agglutinin-like sequence (*ALS*) cell surface protein encoding genes (MacCallum et al., 2009).

To date, *C. albicans* has also been recovered from many different types of both wild and domesticated animals, birds and reptiles (Bougnoux et al., 2004; Buck, 1990; Cafarchia et al., 2006; Edelmann et al., 2005; Jacobsen et al., 2008a; Odds, 1988; Pressler et al., 2003; Tavanti et al., 2005; Wrobel et al., 2008). A number of previous studies have analysed the genetic relationships between *C. albicans* isolates recovered from animal sources and from humans using both Ca3 DNA fingerprinting and MLST (Bougnoux et al., 2004; Edelmann et al., 2005; Jacobsen et al., 2008a; Tavanti et al., 2005; Wrobel et al., 2008). These studies revealed that genetic separation is evident between *C. albicans* isolates recovered from animals and humans, although companion animals such as cats and dogs can yield isolates with similar DSTs to humans (Wrobel et al., 2008). Studies using MLST analysis have suggested that MLST clade 8 is enriched with isolates from wildlife, whereas MLST clade 1 is devoid of such isolates, with the exception of those recovered from primates (Jacobsen et al., 2008a; Wrobel et al., 2008). It has been suggested that MLST clade 1 isolates may be better adapted for colonisation and infection in humans, as isolates belonging to this clade are most frequently recovered from humans (Jacobsen et al., 2008a) and are typically recovered from superficial *C. albicans* infections (Odds et al., 2007). To date, there are 2244 isolate profiles in the *C. albicans* MLST database (date accessed 20.08.13) from isolates which have been recovered in multiple locations throughout the world (Fig. 1). These 2244 isolates comprise 2118 DSTs and to date, can be divided into 18 distinct clades. The *C. albicans* MLST database is arguably the most useful tool for the epidemiological and population analysis of this yeast species, however, the database is only as good as the data submitted to the curator for inclusion. It is important that researchers continue to submit new and previously identified DST and isolate data to the database in order to further enable comparative studies to be undertaken on a large scale. Isolates from large areas of the world are currently lacking or are underrepresented in the database (Fig. 1). The inclusion of isolates from these areas would undoubtedly reveal additional clades and enable a more accurate global view of the population structure of *C. albicans* to be determined.

4. The genome plasticity of *C. albicans*

The genome of *C. albicans* is highly plastic, and clinical isolates exhibit significant karyotype variability (Chibana et al., 2000). Chromosomal rearrangements are tolerated very well in *C. albicans*, and often occur in response to stresses such as heat shock, host-pathogen

interactions and the presence of antifungal drugs (Bouchonville et al., 2009; Forche et al., 2009; Selmecki et al., 2006). Some examples of *C. albicans* genome plasticity and potential evolutionary pathways are described below.

4.1 The parasexual cycle

Prior to the discovery of a set of genes corresponding to the *Saccharomyces cerevisiae* mating type locus in *C. albicans* by Hull and Johnson (1999), the latter species was believed to lack any type of sexual cycle. In *S. cerevisiae*, the mating type locus regulates the sexual cycle, and its homolog in *C. albicans* is referred to as the mating type-like locus (MTL). Similarly to that of *S. cerevisiae*, there are two MTL idiomorphs (types a and α) that are located on chromosome 5 and encode transcription factors that regulate the mating type characteristics (Hull and Johnson, 1999).

Since its discovery, it has been observed that mating occurs in cells that have phenotypically switched from the normal "white" form to the "opaque" mating competent form. The white-opaque switch occurs rarely, but can occur at $37^{o}C$ in anaerobic conditions (Dumitru et al., 2007; Ramirez-Zavala et al., 2008; Whiteway, 2009) and has been shown to occur readily on the skin of mice at 31.5°C (Lachke et al., 2003). Mating-competent strains are naturally homozygous at the MTL locus, displaying either a/a or α/α genotypes following homozygosis of the MTL locus by chromosome loss and subsequent duplication (Berman and Hadany, 2012; Hull and Johnson, 1999; Lockhart et al., 2003; Wu et al., 2005). Pendrak et al. (2004) revealed how deletion of a single *HRB1* allele encoding a haemoglobin response gene could enable white-opaque switching and mating competence in isolates that were heterozygous at the MTL locus. The authors concluded that Hrb1 is involved in a host factor-regulated signaling pathway that controls white-opaque switching and mating in the absence of allelic deletion at the MTL locus.

Cells of the opposite homozygous mating type are able to mate (provided that both are in close proximity to each other and are in the "opaque" form) by conjugation to form tetraploid zygotes, which subsequently undergo concerted chromosome loss until they reach a near-diploid state with high levels of homozygosity and high frequencies of aneuploidy. This is known as the parasexual cycle due to the absence of meiosis, and is thought to occur rarely, possibly only under stressful environmental conditions.

The main function of the parasexual cycle is thought to enable diversification during times of stress, revealing new combinations of recessive traits by loss of heterozygosity (LOH), or resulting in aneuploidy and copy number variation enabling adaptation to adverse environmental conditions. Aneuploidy and revelation of recessive alleles may adversely alter the fitness of the organism, but in highly stressful conditions the parasexual cycle may be a

significant source of diversity permitting adaptation and survival of the organism (Forche et al., 2008; Berman and Hadany, 2012).

4.2 Heterozygosity and haplotype analysis

The predominantly diploid state of *C. albicans* increases genetic variability and therefore the discriminatory power of DNA-sequence based typing methods such as MLST, as each set of sequence data can present homozygous or heterozygous states.

The genome sequence of *C. albicans* is highly heterozygous, with approximately 4% of the 16 Mb genome exhibiting heterozygosity (Jones et al., 2004; van het Hoog et al., 2007). Heterozygosity masks any recessive deleterious mutations that may be present in the genome, and may contribute significantly to strain fitness. Heterozygosity may be lost by mitotic recombination, gene conversion between homologous chromosomes, DNA crossovers or by chromosome loss and duplication. It can affect short tracts of the genome *via* a gene conversion-like process but more frequently, it occurs by DNA crossovers that affect large portions of chromosomes by break induced replication or reciprocal recombination (Forche et al., 2011). Adverse environmental conditions such as exposure to antifungal agents, ultraviolet light, or oxidative stress trigger LOH events in *C. albicans*, which occur more frequently than point mutations (Forche et al., 2009). One study that investigated 12 *C. albicans* clinical isolates that were homozygous at the MTL locus revealed that 10/12 isolates had undergone extensive karyotypic rearrangement, suggesting that homozygosis of the MTL locus could be associated with large scale genome alterations. Legrand et al. (2004) suggested that hyper-recombination occurs in isolates during infection leading to MTL homozygosity and drug resistance following the observation that 10/12 MTL homozygous isolates had undergone extensive karyotypic rearrangements. Other researchers compared growth rates, phenotypes and recombination events in *C. albicans* cultures grown *in vivo* and *in vitro*. This study identified short- and long- range LOH events occurring more frequently in *C. albicans* cultures grown *in vivo* in a murine infection model, than *in vitro* in liquid culture, although cultures incubated *in vivo* grew more slowly compared to *in vitro* cultures (Forche et al., 2009). Exposure to fluconazole has been shown to dramatically increase LOH (Forche et al., 2011) leading to azole-resistance. Homozygosis of the *TAC1* transcription factor-encoding gene on chromosome 5 leads to *TAC1* hyperactivity, which in turn drives high expression of the CDR1 and CDR2 efflux pumps (Coste et al. 2007) leading to resistance to azole drugs including fluconazole, itraconazole and ketoconazole. Interestingly, the *TAC1* gene is located in close proximity to the MTL locus on chromosome 5, correlating with homozygosis of the MTL locus, the first step of the parasexual cycle, which also occurs more frequently in *C.*

11

albicans on exposure to fluconazole. Similarly, LOH in the *FUR1* gene confers resistance to 5FC, and LOH in the *ERG11* gene that encodes lanosterol 14α-demethylase, the target of azoles, is also an azole resistance mechanism (Coste et al., 2007).

Tracing the levels of heterozygosity and particularly LOH throughout populations of *C. albicans* isolates via haplotype analysis has been used to infer several insights regarding the population structure and methods of reproduction of the species. Haplotypes are combinations of adjacent alleles on a chromosome that are inherited together in the absence of any recombination events. Haplotype analysis can be used to investigate recombination events occurring throughout the population and to infer predominant methods of reproduction. The incidence of mating amongst populations of *C. albicans* has been investigated using the SNPs amongst the *C. albicans* MLST loci to define haplotypes. This is achieved by first defining unambiguous haplotypes (homozygous at all polymorphic sites or are heterozygous at only one site) and then by defining ambiguous haplotypes (that are heterozygous at multiple polymorphic sites) according to the theory that they are the same as, or very closely related to, the previously defined unambiguous haplotypes. Haplotype analyses of *C. albicans* isolates have demonstrated that although the method of reproduction is predominantly clonal, recombination events suggestive of mating do happen (Tavanti et al., 2004), supporting the theory that parasexual events can occur, albeit rarely. An analysis of SNPs amongst MLST loci in a *C. albicans* population by Bougnoux et al. revealed excesses of heterozygosity and significant linkage disequilibrium within the five MLST clades examined (clades 1-4 and 11) (Bougnoux et al., 2008b). The authors suggested that mating is a rare occurrence in isolates within the same clades as well as between isolates belonging to different clades. Haplotype maps have been constructed based on the skewed allelic ratios observed in the mitotic progeny of the *C. albicans* laboratory strain SC5314 following whole chromosome aneuploidy (Legrand et al., 2008). The same research group used comparative genomic hybridisation microarrays to infer SNP haplotypes and to track LOH and copy number changes in laboratory derived strains (Abbey et al., 2011). More recently, a complete haplotype of the SC5314 strain has been determined using whole genome sequencing techniques and direct sequencing to resolve the allelic distribution of heterozygous nucleotides (Muzzey et al., 2013).

Recently, Hickman et al. (2013) carried out an experiment tracking LOH events amongst multiple independent loci. This was done using a *GAL1*/Δ*gal1*::*URA3* derivative of the *C. albicans* strain SC5314 that had been marked with a heterozygous counter-selective *GAL1* marker that enabled selection of isolates in which LOH of the *GAL1* gene had occurred, as well as in 123 SNPs located approximately 100 kb apart across most of the genome.

During this study, Hickman et al. observed the absence of heterozygosity for all markers and SNPs examined in one strain (Hickman et al., 2013). Flow cytometry indicated that this strain contained half the DNA content of the diploid control and was in fact haploid. Further flow cytometry detected additional haploid strains following stressful conditions *in vitro* and from *in vivo* mouse models of candidaemia and candidiasis (Hickman et al., 2013). Following prolonged propagation these haploids were capable of auto-diploidisation, resulting in mixed colonies composed of haploid and auto-diploid cells. Furthermore, these haploid strains were viable and mating competent, yielding mating progeny that were reported to grow significantly faster than their haploid parents, although more slowly than the highly heterozygous diploid strain SC5314. The reduced fitness of the haploid and auto-diploid strains is thought to be a consequence of uncovering harmful or disadvantageous recessive mutations that are not lethal (Gow, 2013). Hickman et al. (2013) surmised that restoration of the heterozygous diploid state restored fitness to cells by complementation of recessive alleles, and that haploidisation might be an evolutionary method of removing recessive lethal mutations from *C. albicans* populations.

4.3 Aneuploidy

Aneuploidy refers to the presence of an abnormal number of chromosomes and is an integral part of the parasexual cycle in *C. albicans* as tetraploid progeny result from mating that subsequently undergo concerted chromosome loss. It has been shown that tetraploids grow more slowly and are less virulent than diploid strains (Ibrahim et al., 2005). Tetraploids are however, capable of causing morbidity in a murine infection model, albeit more slowly than diploid strains (Ibrahim et al., 2005). Changes in ploidy have also been detected in cells passaged through a murine infection model, suggesting these may occur during the infection process (Chen et al., 2004; Ibrahim et al., 2005). Aneuploidy may arise due to defects in DNA replication or division machinery and can involve complete or partial chromosomes, and has been observed for each of the eight *C. albicans* chromosomes. Aneuploidy can be detected using array-based comparative genome hydridisation or SNP analysis (Selmecki et al., 2005; Abbey et al., 2011).

Supernumary chromosomes (SNCs) are additional chromosomes that result from extra copies of chromosomes or chromosomal fragments. These chromosomes are unstable and are superfluous containing dispensable information, and vary in size relative to their parent chromosome(s). Selmecki et al. (2009) observed the acquisition of an SNC (composed of two copies of the left arm of chromosome 5 and the right arm of chromosome 3) in later *C. albicans* populations of culture lineages following exposure to fluconazole. This SNC

increased copy numbers of the *ERG11* and *TAC1* genes, both of which can independently and additively confer fluconazole resistance. Whilst beneficial, the presence of this SNC slowed cell growth. Generally, aneuploidy results in reduced cell fitness, although under conditions of increased temperatures, starvation, or exposure to antifungal agents, aneuploidy can confer fitness advantage (Barton and Gull, 1992; Janbon et al., 1998; Legrand et al., 2004). Aneuploidies of chromosomes 2, 5 or 6 have been shown to enable growth on specific carbon sources such as L-sorbose or D-arabinose (Janbon et al., 1998; Rustchenko et al., 1994), and trisomies of chromosomes 3 and 4 are the most common aneuploidal causes of fluconazole resistance (Ibrahim et al., 2005; Perepnikhatka et al., 1999). Work by separate researchers has detected aneuploidy in 50% of *C. albicans* cells that have developed fluconazole resistance (Selmecki et al., 2006; Selmecki et al., 2009).

4.4 Repetitive DNA sequences

The *C. albicans* genome contains a number of different types of repetitive DNA elements that are frequently associated with chromosomal rearrangements. Much of the karyotypic variability of *C. albicans* can be attributed to the major repeat sequence (MRS) regions, the largest nontelomeric homologous sequences that have been identified in *C. albicans* and which account for approximately 3% of the total genomic content (Lephart et al., 2005). Complete MRS regions have been identified in seven of the eight *C. albicans* chromosomes, whereas only a partial MRS region is present on chromosome 3. These MRS recombinational hotspots are made up of three distinct subregions known as the RB2, RPS and HOK domains (Chibana et al., 2000; Chibana et al., 1994). The RPS subregion consists of tandemly repeated "*alt*" units that are 172 bp in length and are localised to a limited region of approximately 100 kb (Chibana et al., 1994). The RPS region serves as a breakpoint for chromosomal rearrangements such as chromosomal length polymorphisms, reciprocal translocations, chromosomal deletions and trisomies of individual chromosomes (Chibana et al., 2000). In the absence of a meiotic cycle, the MRS serves as a source of homology across the majority of the *C. albicans* chromosomes, enabling reciprocal recombination events to occur between non-homologous chromosomes. These events significantly increase genomic diversity and can affect the phenotypes of the resultant cells (Chibana et al., 2000; Lephart and Magee, 2006). The MRS regions have also been identified in seven of the eight chromosomes of the closest relative of *C. albicans*, *C. dubliniensis* (Magee et al., 2008). In *C. dubliniensis* isolates, chromosome R lacks an MRS. Typically, the RPS subregion of *C. dubliniensis* isolates contains higher number of subunits than *C. albicans*, which is considered

the reason for the greater karyotypic variability observed in *C. dubliniensis* (Sullivan et al., 1995; Joly et al., 2002; Magee et al., 2008).

Goodwin and Poulter (2000) observed repetitive regions exhibiting sequence homology to the *C. albicans CTA2* transcription factor in the sub-telomeric regions of several *C. albicans* chromosomes. Van het Hoog et al. (2007) subsequently named this family of sequences as the telomere-associated (*TLO*) genes and concluded that the *TLO* genes are located within 14 kb of the end of each chromosome and orientated in the 5'- 3' direction towards the centromere. A comparative genomic analysis between *C. albicans* and *C. dubliniensis* revealed that the *TLO* gene family has undergone a unique expansion in *C. albicans* since its divergence from its ancestor approximately 20 million years ago (MYA). The *TLO* genes in *C. dubliniensis* exhibit greater amino acid sequence homology with that of the ancestral *C. tropicalis TLO* locus than with those of *C. albicans* (Jackson et al., 2009; Mishra et al., 2007). In contrast to *C. albicans*, most other *Candida* species possess only one *TLO* gene, although *C. dubliniensis* possesses two *TLO* genes, due to gene duplication of the ancestral locus (Jackson et al., 2009). The mechanisms through which the *C. albicans TLO* gene family expansion arose are unclear, but it has been proposed that it might be due to recombination events amongst the highly homologous DNA sequences present between the centromere and parental *TLO* genes (Anderson et al., 2012). The selective advantage of the *TLO* gene family expansion remains to be elucidated, but in *C. dubliniensis*, deletion of the *CdTLO1* gene results in a dramatic reduction in hyphal formation in the presence of serum. Complementation of this *CdTLO1* knockout with a copy of the *C. albicans TLO11* or *TLO12* gene completely restores hyphal production, suggesting that the *TLO* gene family may be involved in the regulation of hyphal morphogenesis (Jackson et al., 2009).

Transposable elements comprise a significant portion of eukaryotic genomes. These can be divided into two main classes, retrotransposons that replicate *via* an RNA intermediate and DNA transposons. Retrotransposons can be subdivided into those that contain long terminal repeats (LTRs), and those that do not (non-LTRs). Five families of retrotransposons have been identified in *S. cerevisiae*, all of which were of the LTR class. In contrast, considerably more families have been identified in *C. albicans*, including non-LTR retrotransposons and DNA transposons (Goodwin and Poulter, 2000), that have not been identified in *S. cerevisiae.* Furthermore, the majority of the retrotransposon population in *C. albicans* appears to be non-functional and of low copy number (Goodwin and Poulter, 2000). Interestingly, researchers have demonstrated how stressful environments can induce retrotransposon activity (Anaya and Roncero, 1996; Wessler, 1996). It is possible that environmental stresses on *C. albicans* has contributed to its diverse retrotransposon

population, in contrast to *S. cerevisiae*, which is indicative of differences in evolutionary pathways between the two species.

5. Evolution of *C. albicans*

5.1 The ancestry of *C. albicans*

The currently understood position of the genus *Candida* within the fungal kingdom is shown in Fig. 3. There are three subphyla in the phylum Ascomycota, one of which is the sub-phylum Saccharomycotina (Fig. 3). The Saccharomycotina sub-phylum contains the class Saccharomycetes and the order Saccharomycetales (Suh et al., 2006). This order consists of approximately 16 families, two of which diverged approximately 170 MYA (Massey et al., 2003; Wolfe and Shields, 1997). One of these families, the *Saccharomycetales incertae sedis* family, contains a subgroup known as the CTG clade, the members of which unusually translate the CTG codon as serine instead of leucine, and can be further distinguished depending on sexual or cryptic cycles. This clade contains the majority of the medically relevant *Candida* species (Fig. 3). Another clade within the order Saccharomycetales consists predominantly of the *Saccharomyces* and consists of species for which the genomes have undergone complete duplication (Diezmann et al., 2004; Fitzpatrick et al., 2006; Wolfe and Shields, 1997) and is referred to as the whole genome duplication (WGD) clade (Fig. 3).

Prior to the widespread availability and application of whole genome sequencing to fungal genomes, phylogenetic and taxonomic analysis of yeasts and moulds were undertaken by morphological analysis, or by single or multigene DNA sequence analysis. Similarly to MLST, DNA sequences from multigene studies were concatenated to produce one sequence for subsequent phylogenetic analysis (Diezmann et al., 2004; Fitzpatrick et al., 2006; Lott et al., 2005; Mishra et al., 2007). It is important to note that these studies were limited by the numbers of species and genes analysed in each investigation.

A comprehensive multigenic phylogenetic study of the order Saccharomycetales examined the DNA sequences of six nuclear genes amongst 38 different species comprising environmental as well as clinical isolates and resolved three main clades (Diezmann et al., 2004). Two years later, a phylogenetic analysis was undertaken on 42 publicly available fungal genomes using both concatenated alignments of 153 universally distributed orthologs, and consensus supertrees, which were constructed using several input trees as datasets. This study showed that both the concatenated alignment and supertree methods are largely congruent, and that the phylogeny depicted by these two methods largely correlated with previous phylogenetic analyses based on single and multigenic datasets, as well as morphological analysis (Fitzpatrick et al., 2006). This suggests phylogenetic resolution of the

fungal kingdom is quite robust, although resolution will improve upon inclusion of more fungal species following whole genome sequencing. The phylogenetic analysis undertaken by Fitzpatrick et al. (2006) illustrates the divergence of *C. albicans* and *C. dubliniensis* from their common ancestor *C. tropicalis* (Fig. 3). Previously, separate researchers studied SNPs throughout 12 housekeeping loci amongst 20 *C. albicans* isolates representative of clades I, II and III defined by DNA fingerprinting using the complex probe Ca3, as well as 10 *C. albicans* isolates from Africa (Lott et al., 2005). These researchers estimated that the divergence of *C. albicans* from its ancestor occurred between 3 and 16 MYA. A separate study investigated the phylogenetics of *C. albicans* based on DNA sequence analysis of the two unlinked tubulin-encoding genes *TUB1* and *TUB2* (located on chromosomes R and 1, respectively) and indicated that *C. albicans* and its sister species *C. dubliniensis* diverged from their ancestral species *C. tropicalis* approximately 20 MYA (Mishra et al., 2007).

5.2 The CTG clade

As mentioned in section 5.1 above, *C. albicans* and several other *Candida* species belongs to the CTG clade which is characterised by alternate codon usage by the species in the clade. The species in this clade translate typically leucine-encoding CUG codons as serine residues. The appearance of the tDNA gene encoding the Ser-tRNA$_{CAG}$ first appeared approximately 270 MYA. This novel Ser-tRNA$_{CAG}$ competed with the wild type Leu-tRNA$_{CAG}$ for the 30,000 CUG codons in the ancestor, and following the divergence of the *Saccharomycetaceae* (Fig. 3), which is thought to have occurred 170 MYA (Gomes et al., 2012; Massey et al., 2003; Miranda et al., 2009) the novel Ser-tRNA$_{CAG}$ was lost from the *Saccharomyces* lineage, which instead retained the standard Leu-tRNA$_{CAG}$. The *Candida* lineage retained the Ser-tRNA$_{CAG}$ reassignment, although CUG codon usage amongst the *Candida* species is quite rare, only 1-2% of the original CUG codons remain (Butler et al., 2009; Massey et al., 2003). The mechanism(s) of this codon reassignment is as yet, unclear. The appearance of the Ser-tRNA$_{CAG}$ prior to codon reassignment supports the codon capture theory which describes neutral codon bias due to genomic G + C content. However, the ambiguous CTG decoding by Ser-tRNA$_{CAG}$ and Leu-tRNA$_{CAG}$ in some *Candida* species (Suzuki et al., 1997) supports the ambiguous intermediate theory, which proposes that selective pressure ultimately drives codon reassignment. Current opinion compromises between the codon capture and ambiguous intermediate theories for codon replacement, in that loss or deletion of a tRNA gene is usually accompanied by the gain of a new tRNA gene for the reassigned codon (Sengupta and Higgs, 2005).

A previous study reconstructed the early effects of genetic code alteration by reintroducing the *S. cerevisiae* Leu-tRNA$_{CAG}$ into a *C. albicans* strain (Miranda et al., 2007) and revealed that this alteration acted as an evolutionary accelerator and generator of phenotypic diversity. Transformation and expression of the *S. cerevisiae* Leu-tRNA$_{CAG}$ in *C. albicans* increased decoding errors, but this was well tolerated, did not significantly slow growth rate and induced morphogenesis, phenotypic switching, increased cell adhesion and the production of secreted aspartyl proteinases (SAPs) and phospholipases (Gomes et al., 2007; Miranda et al., 2007). These studies highlighted how the CTG genetic code alteration was a highly significant event in the evolutionary pathway of the *Candida* species.

5.3 Gene family expansions and evolution of pathogenicity

Candida albicans remains the most pathogenic of the *Candida* species despite the increasing incidence of candidiasis caused by non-*C. albicans Candida* species. Comparative genomic studies using DNA microarrays and whole genome sequence analysis have identified several gene families that have been expanded in pathogenic *Candida* species in contrast to non-pathogenic species (Butler et al., 2009; Jackson et al., 2009; Moran et al., 2004). Whole genome sequence comparisons of seven members of the CTG clade and nine members of the WGD clade identified the enrichment of 21 gene families in pathogenic *Candida* species, and five of these families were particularly enriched in the most pathogenic species (Butler et al., 2009). Three of these five families are associated with the cell wall, those encoding Als adhesins, Hyr1/Iff proteins and the glycosylphosphatidylinositol (GPI) Pga30-like proteins.

There are eight members of the gene family encoding Als glycoprotein adhesins in *C. albicans*, and six in its significantly less virulent sister species *C. dubliniensis*. Each *ALS* gene is composed of a relatively well-conserved 5' domain, a central domain composed of variable numbers of a 108 bp motif and a variable 3' domain. The central domain gives great allelic diversity to these genes, most commonly by variation in the numbers of tandem repeats. Significant clade-specific differences in the numbers of tandem repeats has been observed for the majority of the *C. albicans ALS* genes, as well as the *HYR* genes which are also composed of three main domains, of which the central domain is composed of tandem repeats (MacCallum et al., 2009).

Expansion of these gene families occurs primarily by gene duplication. In *C. albicans*, *ALS5* arose by gene duplication of *ALS1*, and this is mirrored in the *ALS* gene family of *C. dubliniensis* in which Cd36_64800 arose from gene duplication of Cd36_65010, the positional ortholog of the *C. albicans ALS2* gene (Jackson et al., 2009). Gene duplications have also been observed in the secreted aspartyl proteinase (*SAP*) gene family, the *TLO* gene

family (described previously) and the leucine rich *IFA* gene family encoding putative transmembrane proteins. The *TLO* and *IFA* gene families show the largest number discrepancies amongst the *Candida* species. There are 31 *IFA* genes in *C. albicans* in contrast to *C. tropicalis* which has only one. Several of the new *IFA* loci in *C. albicans* have been reported to be a result of gene duplication, however, 6/31 (19.4%) are thought to be non-functional. Interestingly, *C. dubliniensis* has 21 members of the *IFA* gene family, but a larger proportion (14/21, 66.7%) of these are non-functional as a result of pseudogenisation (Jackson et al., 2009).

Direct genomic DNA comparison of the sister species *C. albicans* and *C. dubliniensis* has yielded much information about the enhanced pathogenicity of *C. albicans*. Whilst many families in *C. albicans* have undergone gene expansion, *C. dubliniensis* is in a state of reductive evolution, and has undergone significant pseudogenisation and gene deletion. There is no gene corresponding to the invasin-encoding *ALS3* in *C. dubliniensis* (or in any other *Candida* species), nor is there a gene corresponding to the *C. albicans HYR1* gene (Jackson et al., 2009; Moran et al., 2004). Phylogenetic analysis of some of these gene families has also been used to infer evolutionary pathways of some of these gene families. The solitary ancestral *TLO* gene appears to have undergone a single duplication in *C. dubliniensis*, whereas in *C. albicans* the gene has undergone a more significant expansion. In contrast, the expanded ancestral *IFA* gene family gene has been retained in *C. albicans* but mostly lost from *C. dubliniensis*. Similarly, the ancestral *HYR1* gene has been retained by *C. albicans* but lost from *C. dubliniensis*, the only trace of it in *C. dubliniensis* is a region of 30 Hyr1-homologous amino acid residues identified at the corresponding position to the *HYR1* locus of *C. albicans* (Jackson et al., 2009).

6. Conclusions and future directions

The balance of current opinion indicates that *C. albicans* and its sister species *C. dubliniensis* diverged from their common ancestral species *C. tropicalis* approximately 20 MYA. As evidenced by the identification of homologous recombination, the parasexual cycle, significant LOH, complete or segmental chromosomal aneuploidy and extensive karyotypic variability, the genome of *C. albicans* is highly plastic and the species is very tolerant of gross chromosomal rearrangement. These events appear to occur more frequently when the species is under selective pressure and drive adaptation to less favourable environments. Surprisingly, the sister species *C. dubliniensis* exhibits even higher levels of karyotypic variability than *C. albicans*, yet it appears to be less well able to adapt to unfavourable environments than *C. albicans*. The reason(s) behind this is most likely the evolutionary processes that occurred

during or soon after the divergence of both species from their common ancestor. Whilst *C. dubliniensis* underwent significant gene loss and pseudogenisation, *C. albicans* expanded gene families that appear to be important in virulence.

Despite the genome plasticity of *C. albicans*, the population structure of the species is very robust. Similar clade structures have been identified in the population by Ca3 fingerprinting, MLST and MLMT studies. Furthermore, distinct enrichments of clades with isolates exhibiting antifungal resistance, varying levels of phosphatase activity and abilities to grow in 2 M NaCl, recovered from specific continents, superficial *Candida* infections, anatomical origins, different host species (Table 1), as well as clade-specific numbers of tandem repeat numbers in *ALS* and *HYR* genes all point to separate evolutionary pathways of isolates within each clade. This evidence further suggests that the predominant mode of reproduction of *C. albicans* is clonal, and that the parasexual cycle occurs very infrequently, in response to stressful conditions. It is likely that in time more clades will be identified in the population structure of *C. albicans* following the study of further collections of isolates. Epidemiological and population structure analysis of these isolates will be very important in the further identification of properties that are relatively clade-specific or enriched. For example, clade 1 isolates are more often associated with superficial skin infections, but in time more clades may emerge that are more commonly associated with invasive *Candida* infections.

The application of whole genome sequencing has provided a wealth of information regarding the evolutionary pathway and phylogeny of the pathogenic yeasts, as well as enabling extensive comparisons of the gene families involved in virulence to be made. Further whole genome sequencing of multiple *C. albicans* isolates from different clades, geographical locations, anatomical sources and disease states should provide further useful information regarding virulence gene family expansions, genome evolution and molecular epidemiology of this species.

Acknowledgements

References:

Abbey, D., Hickman, M., Gresham, D., Berman, J., 2011. High-Resolution SNP/CGH Microarrays Reveal the Accumulation of Loss of Heterozygosity in Commonly Used *Candida albicans* Strains. G3 (Bethesda, Md.) 1, 523-530.

Abdulrahim, M.H., McManus, B.A., Flint, S.R., Coleman, D.C. 2013. Genotyping *Candida albicans* from *Candida* leukoplakia and non-*Candida* leukoplakia shows no enrichment of multilocus sequence typing clades but enrichment of ABC genotype C in *Candida* leukoplakia. PLoS One 8(9), e73738.

Anaya, N., Roncero, M.I., 1996. Stress-induced rearrangement of *Fusarium* retrotransposon sequences. Mol Gen Genet 253, 89-94.

Anderson, M.Z., Baller, J.A., Dulmage, K., Wigen, L., Berman, J., 2012. The three clades of the telomere-associated *TLO* gene family of *Candida albicans* have different splicing, localization, and expression features. Eukaryot Cell 11, 1268-1275.

Asmundsdóttir, L.R., Erlendsdóttir, H., Agnarsson, B.A., Gottfredsson, M., 2009. The importance of strain variation in virulence of *Candida dubliniensis* and *Candida albicans*: results of a blinded histopathological study of invasive candidiasis. Clin Microbiol Infect 15, 576-585.

Barton, R.C., Gull, K., 1992. Isolation, characterization, and genetic analysis of monosomic, aneuploid mutants of *Candida albicans*. Mol Microbiol 6, 171-177.

Bartie, K.L., Williams, D.W., Wilson, M.J., Potts, A.J., Lewis, M.A., 2001. PCR fingerprinting of *Candida albicans* associated with chronic hyperplastic candidosis and other oral conditions. J Clin Microbiol 39, 4066-4075.

Bennett, R.J., Johnson, A.D., 2003. Completion of a parasexual cycle in *Candida albicans* by induced chromosome loss in tetraploid strains. The EMBO Journal 22, 2505-2515.

Berman, J., Hadany, L., 2012. Does stress induce (para)sex? Implications for *Candida albicans* evolution. Trends Genet 28, 197-203.

Blignaut, E., Pujol, C., Lockhart, S., Joly, S., Soll, D.R., 2002. Ca3 fingerprinting of *Candida albicans* isolates from human immunodeficiency virus-positive and healthy individuals reveals a new clade in South Africa. J Clin Microbiol 40, 826-836.

Blignaut, E., Molepo, J., Pujol, C., Soll, D.R., Pfaller, M.A., 2005. Clade-related amphotericin B resistance among South African *Candida albicans* isolates. Diagn Microbiol Infect Dis 53, 29-31.

Boerlin, P., Boerlin-Petzold, F., Durussel, C., Addo, M., Pagani, J.L., Chave, J.P., Bille, J., 1995. Cluster of oral atypical *Candida albicans* isolates in a group of human immunodeficiency virus-positive drug users. J Clin Microbiol 33, 1129-1135.

Bouchonville, K., Forche, A., Tang, K.E., Selmecki, A., Berman, J., 2009. Aneuploid chromosomes are highly unstable during DNA transformation of *Candida albicans*. Eukaryot Cell 8, 1554-1566.

Bougnoux, M.E., Aanensen, D.M., Morand, S., Theraud, M., Spratt, B.G., d'Enfert, C., 2004. Multilocus sequence typing of *Candida albicans*: strategies, data exchange and applications. Infect Genet Evol 4, 243-252.

Bougnoux, M.E., Diogo, D., Francois, N., Sendid, B., Veirmeire, S., Colombel, J.F., Bouchier, C., Van Kruiningen, H., d'Enfert, C., Poulain, D., 2006. Multilocus sequence typing reveals intrafamilial transmission and microevolutions of *Candida albicans* isolates from the human digestive tract. J Clin Microbiol 44, 1810-1820.

Bougnoux, M.E., Kac, G., Aegerter, P., d'Enfert, C., Fagon, J.Y., 2008a. Candidemia and candiduria in critically ill patients admitted to intensive care units in France: incidence, molecular diversity, management and outcome. Intensive Care Med 34, 292-299.

Bougnoux, M.E., Morand, S., d'Enfert, C., 2002. Usefulness of multilocus sequence typing for characterization of clinical isolates of *Candida albicans*. J Clin Microbiol 40, 1290-1297.

Bougnoux, M.E., Pujol, C., Diogo, D., Bouchier, C., Soll, D.R., d'Enfert, C., 2008b. Mating is rare within as well as between clades of the human pathogen *Candida albicans*. Fungal Genet Biol 45, 221-231.

Bougnoux, M.E., Tavanti, A., Bouchier, C., Gow, N.A., Magnier, A., Davidson, A.D., Maiden, M.C., D'Enfert, C., Odds, F.C., 2003. Collaborative consensus for optimized multilocus sequence typing of *Candida albicans*. J Clin Microbiol 41, 5265-5266.

Buck, J.D., 1990. Isolation of *Candida albicans* and halophilic Vibrio spp. from aquatic birds in Connecticut and Florida. Appl Environ Microbiol 56, 826-828.

Butler, G., Rasmussen, M.D., Lin, M.F., Santos, M.A., Sakthikumar, S., Munro, C.A., Rheinbay, E., Grabherr, M., Forche, A., Reedy, J.L., Agrafioti, I., Arnaud, M.B., Bates, S., Brown, A.J., Brunke, S., Costanzo, M.C., Fitzpatrick, D.A., de Groot, P.W., 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., Skrzypek, M.S., Soll, D., Staggs, R., Stansfield, I., Stumpf, M.P., Sudbery, P.E., Srikantha, T., Zeng, Q., Berman, J., Berriman, M., Heitman, J., Gow, N.A., Lorenz, M.C., Birren, B.W., Kellis, M., Cuomo, C.A., 2009. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature 459, 657-662.

Cafarchia, C., Camarda, A., Romito, D., Campolo, M., Quaglia, N.C., Tullio, D., Otranto, D., 2006. Occurrence of yeasts in cloacae of migratory birds. Mycopathologia 161, 229-234.

Caugant, D.A., Sandven, P., 1993. Epidemiological analysis of *Candida albicans* strains by multilocus enzyme electrophoresis. J Clin Microbiol 31, 215-220.

Chavez-Galarza, J., Pais, C., Sampaio, P., 2010. Microsatellite typing identifies the major clades of the human pathogen *Candida albicans*. Infect Genet Evol 10, 697-702.

Chen, K.W., Chen, Y.C., Lo, H.J., Odds, F.C., Wang, T.H., Lin, C.Y., Li, S.Y., 2006. Multilocus sequence typing for analyses of clonality of *Candida albicans* strains in Taiwan. J Clin Microbiol 44, 2172-2178.

Chen, X., Magee, B.B., Dawson, D., Magee, P.T., Kumamoto, C.A., 2004. Chromosome 1 trisomy compromises the virulence of *Candida albicans*. Mol Microbiol 51, 551-565.

Chibana, H., Beckerman, J.L., Magee, P.T., 2000. Fine-resolution physical mapping of genomic diversity in *Candida albicans*. Genome Res 10, 1865-1877.

Chibana, H., Iwaguchi, S., Homma, M., Chindamporn, A., Nakagawa, Y., Tanaka, K., 1994. Diversity of tandemly repetitive sequences due to short periodic repetitions in the chromosomes of *Candida albicans*. J Bacteriol 176, 3851-3858.

Cliff, P.R., Sandoe, J.A., Heritage, J., Barton, R.C., 2008. Use of multilocus sequence typing for the investigation of colonisation by *Candida albicans* in intensive care unit patients. J Hosp Infect 69, 24-32.

Coste, A., Selmecki, A., Forche, A., Diogo, D., Bougnoux, M.E., d'Enfert, C., Berman, J., Sanglard, D., 2007. Genotypic evolution of azole resistance mechanisms in sequential *Candida albicans* isolates. Eukaryot Cell 6, 1889-1904.

Da Matta, D.A., Melo, A.S., Guimaraes, T., Frade, J.P., Lott, T.J., Colombo, A.L., 2010. Multilocus sequence typing of sequential *Candida albicans* isolates from patients with persistent or recurrent fungemia. Med Mycol 48, 757-762.

Dalle, F., Franco, N., Lopez, J., Vagner, O., Caillot, D., Chavanet, P., Cuisenier, B., Aho, S., Lizard, S., Bonnin, A., 2000. Comparative genotyping of *Candida albicans* bloodstream and nonbloodstream isolates at a polymorphic microsatellite locus. J Clin Microbiol 38, 4554-4559.

Diezmann, S., Cox, C.J., Schonian, G., Vilgalys, R.J., Mitchell, T.G., 2004. Phylogeny and evolution of medical species of *Candida* and related taxa: a multigenic analysis. J Clin Microbiol 42, 5624-5635.

Dodgson, A.R., Dodgson, K.J., Pujol, C., Pfaller, M.A., Soll, D.R., 2004. Clade-specific flucytosine resistance is due to a single nucleotide change in the *FUR1* gene of *Candida albicans*. Antimicrob Agents Chemother 48, 2223-2227.

Dumitru, R., Navarathna, D.H., Semighini, C.P., Elowsky, C.G., Dumitru, R.V., Dignard, D., Whiteway, M., Atkin, A.L., Nickerson, K.W., 2007. *In vivo* and *in vitro* anaerobic mating in *Candida albicans*. Eukaryot Cell 6, 465-472.

Edelmann, A., Kruger, M., Schmid, J., 2005. Genetic relationship between human and animal isolates of *Candida albicans*. J Clin Microbiol 43, 6164-6166.

Feil, E.J., Enright, M.C., 2004. Analyses of clonality and the evolution of bacterial pathogens. Curr Opin Microbiol 7, 308-313.

Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P., Spratt, B.G., 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186, 1518-1530.

Fitzpatrick, D.A., Logue, M.E., Stajich, J.E., Butler, G., 2006. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol Biol 6, 99.

Forche, A., Abbey, D., Pisithkul, T., Weinzierl, M.A., Ringstrom, T., Bruck, D., Petersen, K., Berman, J., 2011. Stress alters rates and types of loss of heterozygosity in *Candida albicans*. mBio 2(4):e00129-11.

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

Forche, A., Magee, P.T., Selmecki, A., Berman, J., May, G., 2009. Evolution in *Candida albicans* populations during a single passage through a mouse host. Genetics 182, 799-811.

Forche, A., Schonian, G., Graser, Y., Vilgalys, R., Mitchell, T.G., 1999. Genetic structure of typical and atypical populations of *Candida albicans* from Africa. Fungal Genet Biol 28, 107-125.

Fundyga, R.E., Lott, T.J., Arnold, J., 2002. Population structure of *Candida albicans*, a member of the human flora, as determined by microsatellite loci. Infect Genet Evol 2, 57-68.

Gammelsrud, K.W., Lindstad, B.L., Gaustad, P., Ingebretsen, A., Hoiby, E.A., Brandtzaeg, P., Sandven, P., 2012. Multilocus sequence typing of serial *Candida albicans* isolates from children with cancer, children with cystic fibrosis and healthy controls. Med Mycol 50, 619-626.

Ge, S.H., Xie, J., Xu, J., Li, J., Li, D.M., Zong, L.L., Zheng, Y.C., Bai, F.Y., 2012. Prevalence of specific and phylogenetically closely related genotypes in the population of *Candida albicans* associated with genital candidiasis in China. Fungal Genet Biol 49, 86-93.

Gil-Lamaignere, C., Roilides, E., Hacker, J., Muller, F.M., 2003. Molecular typing for fungi--a critical review of the possibilities and limitations of currently and future methods. Clin Microbiol Infect 9, 172-185.

Gomes, A.C., Miranda, I., Silva, R.M., Moura, G.R., Thomas, B., Akoulitchev, A., Santos, M.A., 2007. A genetic code alteration generates a proteome of high diversity in the human pathogen *Candida albicans*. Genome Biol 8, R206.

Gomes, A.C., Moura, G.R., Santos, M.A.S., 2012. The Genetic Code of the *Candida* CTG clade, in: Calderone, R.A., Clancy, C.J. (Eds.), Candida and candidiasis, 2nd ed. ASM Press, United States of America, pp. 45-55.

Gong, Y.B., Zheng, J.L., Jin, B., Zhuo, D.X., Huang, Z.Q., Qi, H., Zhang, W., Duan, W., Fu, J.T., Wang, C.J., Mao, Z.B., 2012. Particular *Candida albicans* strains in the digestive tract of dyspeptic patients, identified by multilocus sequence typing. PloS One 7, e35311.

Goodwin, T.J., Poulter, R.T., 2000. Multiple LTR-retrotransposon families in the asexual yeast *Candida albicans*. Genome Res 10, 174-191.

Gow, N.A., 2013. Fungal biology: Multiple mating strategies. Nature 494, 45-46.

Gräser, Y., Volovsek, M., Arrington, J., Schonian, G., Presber, W., Mitchell, T.G., Vilgalys, R., 1996. Molecular markers reveal that population structure of the human pathogen *Candida albicans* exhibits both clonality and recombination. Proc Natl Acad Sci U S A. 93, 12473-12477.

Hickman, M.A., Zeng, G., Forche, A., Hirakawa, M.P., Abbey, D., Harrison, B.D., Wang, Y.M., Su, C.H., Bennett, R.J., Wang, Y., Berman, J., 2013. The 'obligate diploid' *Candida albicans* forms mating-competent haploids. Nature 494, 55-59.

Hull, C.M., Johnson, A.D., 1999. Identification of a mating type-like locus in the asexual pathogenic yeast *Candida albicans*. Science 285, 1271-1275.

Ibrahim, A.S., Magee, B.B., Sheppard, D.C., Yang, M., Kauffman, S., Becker, J., Edwards, J.E., Jr., Magee, P.T., 2005. Effects of ploidy and mating type on virulence of *Candida albicans*. Infect Immun 73, 7366-7374.

Jackson, A.P., Gamble, J.A., Yeomans, T., Moran, G.P., Saunders, D., Harris, D., Aslett, M., Barrell, J.F., Butler, G., Citiulo, F., Coleman, D.C., de Groot, P.W., Goodwin, T.J., Quail, M.A., McQuillan, J., Munro, C.A., Pain, A., Poulter, R.T., Rajandream, M.A., Renauld, H., Spiering, M.J., Tivey, A., Gow, N.A., Barrell, B., Sullivan, D.J., Berriman, M., 2009. Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. Genome Res 19, 2231-2244.

Jacobsen, M.D., Bougnoux, M.E., d'Enfert, C., Odds, F.C., 2008a. Multilocus sequence typing of *Candida albicans* isolates from animals. Res Microbiol 159, 436-440.

Jacobsen, M.D., Rattray, A.M., Gow, N.A., Odds, F.C., Shaw, D.J., 2008b. Mitochondrial haplotypes and recombination in *Candida albicans*. Med Mycol 46, 647-654

Janbon, G., Sherman, F., Rustchenko, E., 1998. Monosomy of a specific chromosome determines L-sorbose utilization: a novel regulatory mechanism in *Candida albicans*. Proceedings of the National Academy of Sciences of the United States of America 95, 5150-5155.

Joly, S., Pujol, C., Soll, D.R., 2002. Microevolutionary changes and chromosomal translocations are more frequent at RPS loci in *Candida dubliniensis* than in *Candida albicans*. Infect Genet Evol 2, 19-37.

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., Scherer, S., 2004. The diploid genome sequence of *Candida albicans*. Proc Natl Acad Sci USA 101, 7329-7334.

Kumamoto, C.A., 2011. Inflammation and gastrointestinal *Candida* colonization. Curr Opin Microbiol 14, 386-391.

Lachke, S.A., Lockhart, S.R., Daniels, K.J., Soll, D.R., 2003. Skin facilitates *Candida albicans* mating. Infect Immun 71, 4970-4976.

Legrand, M., Lephart, P., Forche, A., Mueller, F.M., Walsh, T., Magee, P.T., Magee, B.B., 2004. Homozygosity at the MTL locus in clinical strains of *Candida albicans*: karyotypic rearrangements and tetraploid formation. Mol Microbiol 52, 1451-1462.

Legrand, M., Forche, A., Selmecki, A., Chan, C., Kirkpatrick, D.T., Berman, J., 2008. Haplotype mapping of a diploid non-meiotic organism using existing and induced aneuploidies. PLoS Genet 4, e1.

Lephart, P.R., Chibana, H., Magee, P.T., 2005. Effect of the major repeat sequence on chromosome loss in *Candida albicans*. Eukaryot Cell 4, 733-741.

Lephart, P.R., Magee, P.T., 2006. Effect of the major repeat sequence on mitotic recombination in *Candida albicans*. Genetics 174, 1737-1744.

Lockhart, S.R., Daniels, K.J., Zhao, R., Wessels, D., Soll, D.R., 2003. Cell biology of mating in *Candida albicans*. Eukaryot Cell 2, 49-61.

Lott, T.J., Fundyga, R.E., Kuykendall, R.J., Arnold, J., 2005. The human commensal yeast, *Candida albicans,* has an ancient origin. Fungal Genet Biol 42, 444-451.

MacCallum, D.M., Castillo, L., Nather, K., Munro, C.A., Brown, A.J., Gow, N.A., Odds, F.C., 2009. Property differences among the four major *Candida albicans* strain clades. Eukaryot Cell 8, 373-387.

Magee, B.B., Sanchez, M.D., Saunders, D., Harris, D., Berriman, M., Magee, P.T., 2008. Extensive chromosome rearrangements distinguish the karyotype of the hypovirulent species *Candida dubliniensis* from the virulent *Candida albicans*. Fungal Genet Biol 45, 338-350.

Massey, S.E., Moura, G., Beltrao, P., Almeida, R., Garey, J.R., Tuite, M.F., Santos, M.A., 2003. Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in *Candida* spp. Genome Res 13, 544-557.

McCullough, M.J., Clemons, K.V., Stevens, D.A., 1999. Molecular and phenotypic characterization of genotypic *Candida albicans* subgroups and comparison with *Candida dubliniensis* and *Candida stellatoidea*. J Clin Microbiol 37, 417-421.

McManus, B.A., Coleman, D.C., Moran, G., Pinjon, E., Diogo, D., Bougnoux, M.E., Borecka-Melkusova, S., Bujdakova, H., Murphy, P., d'Enfert, C., Sullivan, D.J., 2008. Multilocus Sequence Typing Reveals that the Population Structure of *Candida dubliniensis* Is Significantly Less Divergent than That of *Candida albicans*. J Clin Microbiol 46, 652-664.

McManus, B.A., Maguire, R., Cashin, P.J., Claffey, N., Flint, S., Abdulrahim, M.H., Coleman, D.C., 2012. Enrichment of multilocus sequence typing clade 1 with oral *Candida albicans* isolates in patients with untreated periodontitis. J Clin Microbiol 50, 3335-3344.

McManus, B.A., McGovern, E., Moran, G.P., Healy, C.M., Nunn, J., Fleming, P., Costigan, C., Sullivan, D.J., Coleman, D.C., 2011. Microbiological screening of Irish patients with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy reveals persistence of *Candida albicans* strains, gradual reduction in susceptibility to azoles, and incidences of clinical signs of oral candidiasis without culture evidence. J Clin Microbiol 49, 1879-1889.

McManus, B.A., Sullivan, D.J., Moran, G.P., d'Enfert, C., Bougnoux, M.E., Nunn, M.A., Coleman, D.C., 2009. Genetic differences between avian and human isolates of *Candida dubliniensis*. Emerg Infect Dis 15, 1467-1470.

Meyer, W., Lieckfeldt, E., Kuhls, K., Freedman, E.Z., Borner, T., Mitchell, T.G., 1993. DNA- and PCR-fingerprinting in fungi. Exs 67, 311-320.

Miranda, I., Rocha, R., Santos, M.C., Mateus, D.D., Moura, G.R., Carreto, L., Santos, M.A., 2007. A genetic code alteration is a phenotype diversity generator in the human pathogen *Candida albicans*. PloS One 2, e996.

Miranda, T.T., Vianna, C.R., Rodrigues, L., Monteiro, A.S., Rosa, C.A., Correa, A., Jr., 2009. Diversity and frequency of yeasts from the dorsum of the tongue and necrotic root canals associated with primary apical periodontitis. Int Endod J 42, 839-844.

Mishra, P.K., Baum, M., Carbon, J., 2007. Centromere size and position in *Candida albicans* are evolutionarily conserved independent of DNA sequence heterogeneity. Mol Genet Genomics 278, 455-465.

Moran, G., Stokes, C., Thewes, S., Hube, B., Coleman, D.C., Sullivan, D., 2004. Comparative genomics using *Candida albicans* DNA microarrays reveals absence and divergence of virulence-associated genes in *Candida dubliniensis*. Microbiol 150, 3363-3382.

Muzzey, D., Schwartz, K., Weissman, J.S., Sherlock, G., 2013. Assembly of a phased diploid *Candida albicans* genome facilitates allele-specific measurements and provides a simple model for repeat and indel structure. Genome Biol 14, R97.

Odds, F.C., 1988. Ecology of *Candida* and epidemiology of candidosis. in: Odds, F.C. (Ed.), *Candida* and Candidosis. Bailière Tindall, London, pp. 68-92.

Odds, F.C., Bougnoux, M.E., Shaw, D.J., Bain, J.M., Davidson, A.D., Diogo, D., Jacobsen, M.D., Lecomte, M., Li, S.Y., Tavanti, A., Maiden, M.C., Gow, N.A., d'Enfert, C., 2007. Molecular phylogenetics of *Candida albicans*. Eukaryot Cell 6, 1041-1052.

Odds, F.C., Davidson, A.D., Jacobsen, M.D., Tavanti, A., Whyte, J.A., Kibbler, C.C., Ellis, D.H., Maiden, M.C., Shaw, D.J., Gow, N.A., 2006. *Candida albicans* strain maintenance, replacement, and microvariation demonstrated by multilocus sequence typing. J Clin Microbiol 44, 3647-3658.

Odds, F.C., Jacobsen, M.D., 2008. Multilocus sequence typing of pathogenic *Candida* species. Eukaryot Cell 7, 1075-1084.

Pendrak, M.L., Yan, S.S., Roberts, D.D., 2004. Hemoglobin regulates expression of an activator of mating-type locus alpha genes in *Candida albicans*. Eukaryot Cell 3, 764-775.

Perepnikhatka, V., Fischer, F.J., Niimi, M., Baker, R.A., Cannon, R.D., Wang, Y.K., Sherman, F., Rustchenko, E., 1999. Specific chromosome alterations in fluconazole-resistant mutants of *Candida albicans*. J Bacteriol 181, 4041-4049.

Pfaller, M.A., Castanheira, M., Messer, S.A., Moet, G.J., Jones, R.N., 2010. Variation in *Candida spp.* distribution and antifungal resistance rates among bloodstream infection isolates by patient age: report from the SENTRY Antimicrobial Surveillance Program (2008-2009). Diagn Microbiol Infect Dis 68, 278-283.

Pressler, B.M., Vaden, S.L., Lane, I.F., Cowgill, L.D., Dye, J.A., 2003. *Candida* spp. urinary tract infections in 13 dogs and seven cats: predisposing factors, treatment, and outcome. J Am Anim Hosp Assoc 39, 263-270.

Pujol, C., Messer, S.A., Pfaller, M., Soll, D.R., 2003. Drug resistance is not directly affected by mating type locus zygosity in *Candida albicans*. Antimicrob Agents Chemother 47, 1207-1212.

Pujol, C., Pfaller, M., Soll, D.R., 2002. Ca3 fingerprinting of *Candida albicans* bloodstream isolates from the United States, Canada, South America, and Europe reveals a European clade. J Clin Microbiol 40, 2729-2740.

Pujol, C., Joly, S., Lockhart, S.R., Noel, S., Tibayrenc, M., Soll, D.R., 1997. Parity among the randomly amplified polymorphic DNA method, multilocus enzyme electrophoresis, and Southern blot hybridization with the moderately repetitive DNA probe Ca3 for fingerprinting *Candida albicans*. J Clin Microbiol 35, 2348-2358.

Ramirez-Zavala, B., Reuss, O., Park, Y.N., Ohlsen, K., Morschhauser, J., 2008. Environmental induction of white-opaque switching in *Candida albicans*. PLoS Pathog 4, e1000089.

Robles, J.C., Koreen, L., Park, S., Perlin, D.S., 2004. Multilocus sequence typing is a reliable alternative method to DNA fingerprinting for discriminating among strains of *Candida albicans*. J Clin Microbiol 42, 2480-2488.

Rustchenko, E.P., Howard, D.H., Sherman, F., 1994. Chromosomal alterations of *Candida albicans* are associated with the gain and loss of assimilating functions. J Bacteriol 176, 3231-3241.

Saghrouni, F., Ben Abdeljelil, J., Boukadida, J., Ben Said, M., 2013. Molecular methods for strain typing of *Candida albicans*: a review. J Appl Microbiol 14, 1559-1574.

Sampaio, P., Gusmao, L., Alves, C., Pina-Vaz, C., Amorim, A., Pais, C., 2003. Highly polymorphic microsatellite for identification of *Candida albicans* strains. J Clin Microbiol 41, 552-557.

Sampaio, P., Gusmao, L., Correia, A., Alves, C., Rodrigues, A.G., Pina-Vaz, C., Amorim, A., Pais, C., 2005. New microsatellite multiplex PCR for *Candida albicans* strain typing reveals microevolutionary changes. J Clin Microbiol 43, 3869-3876.

Scherer, S., Stevens, D.A., 1988. A *Candida albicans* dispersed, repeated gene family and its epidemiologic applications. Proc Natl Acad Sci USA 85, 1452-1456.

Selmecki, A., Bergmann, S., Berman, J., 2005. Comparative genome hybridization reveals widespread aneuploidy in *Candida albicans* laboratory strains. Mol Microbiol 55, 1553-1565.

Selmecki, A., Forche, A., Berman, J., 2006. Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. Science 313, 367-370.

Selmecki, A.M., Dulmage, K., Cowen, L.E., Anderson, J.B., Berman, J., 2009. Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. PLoS Genet 5, e1000705.

Sengupta, S., Higgs, P.G., 2005. A unified model of codon reassignment in alternative genetic codes. Genetics 170, 831-840.

Shin, J.H., Bougnoux, M.E., d'Enfert, C., Kim, S.H., Moon, C.J., Joo, M.Y., Lee, K., Kim, M.N., Lee, H.S., Shin, M.G., Suh, S.P., Ryang, D.W., 2011. Genetic diversity among Korean *Candida albicans* bloodstream isolates: assessment by multilocus sequence typing and restriction endonuclease analysis of genomic DNA by use of BssHII. J Clin Microbiol 49, 2572-2577.

Soll, D.R., 2000. The ins and outs of DNA fingerprinting the infectious fungi. Clin Microbiol Rev 13, 332-370.

Soll, D.R., Pujol, C., 2003. *Candida albicans* clades. FEMS Immunol Med Microbiol 39, 1-7.

Stephan, F., Bah, M.S., Desterke, C., Rezaiguia-Delclaux, S., Foulet, F., Duvaldestin, P., Bretagne, S., 2002. Molecular diversity and routes of colonization of *Candida albicans* in a surgical intensive care unit, as studied using microsatellite markers. Clin Infect Dis 35, 1477-1483.

Stokes, C., Moran, G.P., Spiering, M.J., Cole, G.T., Coleman, D.C., Sullivan, D.J., 2007. Lower filamentation rates of *Candida dubliniensis* contribute to its lower virulence in comparison with *Candida albicans*. Fungal Genet Biol 44, 920-931.

Suh, S.O., Blackwell, M., Kurtzman, C.P., Lachance, M.A., 2006. Phylogenetics of Saccharomycetales, the ascomycete yeasts. Mycologia 98, 1006-1017.

Sullivan, D.J., Westerneng, T.J., Haynes, K.A., Bennett, D.E., Coleman, D.C., 1995. *Candida dubliniensis sp. nov.*: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals. Microbiol 141, 1507-1521.

Suzuki, T., Ueda, T., Watanabe, K., 1997. The 'polysemous' codon--a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. EMBO J 16, 1122-1134.

Takakura, S., Ichiyama, S., Bain, J.M., Davidson, A.D., Jacobsen, M.D., Shaw, D.J., Gow, N.A., Odds, F.C., 2008. Comparison of *Candida albicans* strain types among isolates from three countries. Int J Med Microbiol 298, 663-668.

Tavanti, A., Davidson, A.D., Fordyce, M.J., Gow, N.A., Maiden, M.C., Odds, F.C., 2005. Population structure and properties of *Candida albicans*, as determined by multilocus sequence typing. J Clin Microbiol 43, 5601-5613.

Tavanti, A., Gow, N.A., Maiden, M.C., Odds, F.C., Shaw, D.J., 2004. Genetic evidence for recombination in *Candida albicans* based on haplotype analysis. Fungal Genet Biol 41, 553-562.

Tavanti, A., Gow, N.A., Senesi, S., Maiden, M.C., Odds, F.C., 2003. Optimization and validation of multilocus sequence typing for *Candida albicans*. J Clin Microbiol 41, 3765-3776.

Thompson, G.R., 3rd, Patel, P.K., Kirkpatrick, W.R., Westbrook, S.D., Berg, D., Erlandsen, J., Redding, S.W., Patterson, T.F., 2010. Oropharyngeal candidiasis in the era of antiretroviral therapy. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 109, 488-495.

van het Hoog, M., Rast, T.J., Martchenko, M., Grindle, S., Dignard, D., Hogues, H., Cuomo, C., Berriman, M., Scherer, S., Magee, B.B., Whiteway, M., Chibana, H., Nantel, A., Magee,

P.T., 2007. Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. Genome Biol 8, R52.

Wessler, S.R., 1996. Turned on by stress. Plant retrotransposons. Curr Biol 6, 959-961.

Whiteway, M., 2009. Yeast mating: putting some fizz into fungal sex? Curr Biol 19, R258-260.

Wolfe, K.H., Shields, D.C., 1997. Molecular evidence for an ancient duplication of the entire yeast genome. Nature 387, 708-713.

Wrobel, L., Whittington, J.K., Pujol, C., Oh, S.H., Ruiz, M.O., Pfaller, M.A., Diekema, D.J., Soll, D.R., Hoyer, L.L., 2008. Molecular phylogenetic analysis of a geographically and temporally matched set of *Candida albicans* isolates from humans and nonmigratory wildlife in central Illinois. Eukaryot Cell 7, 1475-1486.

Wu, W., Pujol, C., Lockhart, S.R., Soll, D.R., 2005. Chromosome loss followed by duplication is the major mechanism of spontaneous mating-type locus homozygosis in *Candida albicans*. Genetics 169, 1311-1327.

Zomorodian, K., Haghighi, N.N., Rajaee, N., Pakshir, K., Tarazooie, B., Vojdani, M., Sedaghat, F., Vosoghi, M., 2011. Assessment of *Candida* species colonization and denture-related stomatitis in complete denture wearers. Med Mycol 49, 208-211.

Web references
http://www.catalogueoflife.org/ (date accessed 21.08.13)
Roskov Y., Kunze T., Paglinawan L., Abucay L., Orrell T., Nicolson D., Culham A., Bailly N., Kirk P., Bourgoin T., Baillargeon G., Hernandez F., De Wever A., Didžiulis V., eds (2013). Species 2000 and ITIS Catalogue of Life, 15th August 2013. Digital resource at www.catalogueoflife.org/col. Species 2000: Reading, UK.

http://www.mycobank.org/ (date accessed 21.08.13)
Crous P.W., Gams W., Stalpers J.A., Robert V., Stegehuis G., 2004. MycoBank: an online initiative to launch mycology into the 21st century. Studies in Mycology 50: 19–22.
Robert V., Stegehuis G., Stalpers J. 2005. The MycoBank engine and related databases. http://www.mycobank.org.

http://calbicans.mlst.net/ (date accessed 20.08.13)
Bougnoux, M.E., Tavanti, A., Bouchier, C., Gow, N.A., Magnier, A., Davidson, A.D., Maiden, M.C., D'Enfert, C., Odds, F.C., 2003. Collaborative consensus for optimized multilocus sequence typing of *Candida albicans*. J Clin Microbiol 41, 5265-5266.

1   **FIG. 1.**

2   **Global distribution of *C. albicans* isolates currently included in the MLST**

3   **database.** Numbers beside each pin indicate the number of isolates currently included

4   in the MLST database from each country (date accessed 20.08.2013). Geographical

5   location information is available for 2083 of the 2244 isolates currently in the MLST

6   database, but is not available for the remaining isolates. Further epidemiological and

7   DST information for these isolates can be retrieved from the *C. albicans* MLST

8   database online (http://calbicans.mlst.net/). This figure highlights the geographical

9   locations that are currently underrepresented in the MLST database; examination of

10   isolates recovered from these locations may reveal the presence of additional MLST

11   clades.

12

13   **FIG. 2.**

14   **Population snapshot of the 2118 *C. albicans* distinct DSTs currently in the MLST**

15   **database (www.calbicans.mlst.net) defined using eBURST.** In the snapshot a single

16   line joins DSTs that differ by only one of the seven loci. The putative founding DST

17   of each clonal complex (CC) and the MLST clade (defined by Odds et al., 2007; Shin

18   et al., 2011) to which this DST belongs is indicated beside each of the larger CCs.

19   Founding DSTs for each CC are indicated in blue, and founders of sub-groups are

20   indicated in yellow. Single black dots represent singletons.

21

22   **FIG. 3. Summary of the current understanding of the ancestry and phylogeny of**

23   ***C. albicans.***

24   The evolutionary pathway of *C. albicans* is indicated in bold italicised typeface on a

25   darker grey background. Taxonomic classifications are indicated in plain typeface on

26   a lighter grey background. The genus *Candida* is currently classified with the

27   *Saccharomycetales incertae sedis* until its family classification is more accurately

28   resolved. The divergence of *C. albicans* and *C. dubliniensis* from its *C. tropicalis*

29   ancestor, thought to have occurred approximately 20 MYA is illustrated, as is the

30   separation of the WGD and CTG lineages. Summary was constructed based on

31   previous fungal phylogenetic studies (Butler et al., 2009; Diezmann et al., 2004;

32   Fitzpatrick et al., 2008; Roskov et al., 2013; Suh et al., 2006), as well as using the

33   online databases http://www.catalogueoflife.org/ and http://www.mycobank.org/.

34

TABLE 1. Features associated with MLST clades in the population structure of *C. albicans*.

Clade[a]	Origin of isolates	ABC genotype[b]	Geographical enrichment[c]	Reduced susceptibility	Date of isolation	Other clade-specific associations	Reference
1	Superficial infection Vaginal infection Oral carriage	A	Worldwide	FLC, 5FC, Terbafine	Pre-1990-2006	Growth in 2 M NaCl	Tavanti et al. (2005) Odds et al. (2007) McManus et al. (2012) Ge et al. (2012) MacCallum et al. (2009) Abdulrahim et al. (2013)
2	Bloodstream Oropharyngeal	A	UK		1990-2006	Lower acid phosphatase activity	Tavanti et al. (2005) Takakura et al. (2008) Odds et al. (2007) MacCallum et al. (2009)
3		B	USA	FLC	1990-1999		Odds et al. (2007) Tavanti et al. (2005) Takakura et al. (2008)
4	Bloodstream	B/C	Middle East/ Africa	Amphotericin B	Pre-1990 and 2000-2006		Blignaut et al. (2005) Odds et al. (2007) Tavanti et al. (2005)
5	Oropharyngeal	B/C	Europe/UK				Odds et al. (2007) Tavanti et al. (2005)
6	Oropharyngeal	B	UK	FLC			Odds et al. (2007) Tavanti et al. (2005)
7		A					Odds et al. (2007) Tavanti et al. (2005)
8	Bloodstream Wildlife animals	A/B	South America				Odds et al. (2007)
9		A					Odds et al. (2007)
10		B	Continental Europe				Odds et al. (2007)
11		A/C	Europe/UK		2000-2006		Odds et al. (2007)
12		B/C					Odds et al. (2007)
13	Vaginal	A	Africa				Odds et al. (2007)
14		B	Asia				Gong et al. (2012) Odds et al. (2007)
15		B/C	Asia				Odds et al. (2007)
16		B	Asia				Odds et al. (2007)
17		A/B	Asia				Odds et al. (2007)
18	Dyspeptic patients	ND	Asia				Gong et al. (2012) Shin et al. (2011)

C6
DST409

C11
DST461

C16
DST669

C11
DST538

C12
DST601

C18
DST727

C8
DST840

C4
DST659

C9
DST918

C10
DST304

C1
DST69

C3
DST344

C1
DST766

C8
DST365

C2
DST155

C4
DST124

| Superkingdom | **Eukaryota** | | | | |

| Kingdom | Protista | **Fungi** | Plantae | Animalia |

| Phylum | Glomeromycota | Chytridiomycota | **Ascomycota** | Basidiomycota | Zygomycota | "Deuteromycetes" |

| Sub-Phylum | Pezizomycotina | **Saccharomycotina** | Taphrinomycotina |

| Class | 18 classes including **Saccharomycetes** |

| Order | **Saccharomycetales** | → 14 additional families |

WGD

13 additional genera ← | Family | Saccharomycetaceae |

| Family | **Saccharomycetales incertae sedis** |

CTG

| Genus | Saccharomyces |

| Genus | **Candida** |

| Species | S. cerevisiae
S. paradoxus
S. bayanus
9 other species |

| Species | **C. tropicalis** |

| Species | ~374 other Candida species including C. parapsilosis C. lusitaniae & C. guilliermondii |

| Species | **C. albicans** | C. dubliniensis |