Published in final edited form as:
Fungal Genet Biol. 2008 March ; 45(3): 221–231.

Mating is Rare Within as Well as Between Clades of the Human Pathogen *Candida albicans*

Marie-Elisabeth Bougnoux[1,2,§], **Claude Pujol**[3,§], **Dorothée Diogo**[1], **Christiane Bouchier**[4], **David R. Soll**[3], and **Christophe d'Enfert**[1,*]

[1]*Unité Biologie et Pathogénicité Fongiques, INRA USC2019 and* [4]*Pasteur Génopole-Ile de France, Département Génomes et Génétique, Institut Pasteur, Paris, France* [2]*Laboratoire de Parasitologie-Mycologie, Service de Microbiologie, Hôpital Necker-Enfants Malades, Université Paris 5-René Descartes, Faculté de Médecine, Paris, France* [3]*Department of Biological Sciences, The University of Iowa, Iowa City, IA, USA*

Abstract

Candida albicans is a diploid yeast that can undergo mating and a parasexual cycle, but is apparently unable to undergo meiosis. Characterization of the population structure of *C. albicans* has shown that reproduction is largely clonal and that mating, if it occurs, is rare or limited to genetically related isolates. Because molecular typing has delineated distinct clades in *C. albicans*, we have tested whether recombination was common within clades, but rare between clades. Two hundred and three *C. albicans* isolates have been subjected to MultiLocus Sequence Typing (MLST) and the haplotypes at heterozygous MLST genotypes characterized. The *C. albicans* isolates were distributed among nine clades, of which five corresponded to those previously identified by Ca3 fingerprinting. In each of these clades with more than 10 isolates, polymorphic nucleotide positions located on between 3 and 4 of the 6 loci were in Hardy-Weinberg disequilibrium. Moreover, each of these polymorphic sites contained excess heterozygotes. This was confirmed by an expanded analysis performed on a recently published MLST dataset for 1,044 isolates. On average, 66% of polymorphic positions in the individual clades were in significant excess of heterozygotes over the five clades. These data indicate that mating within clades as well as self-fertilization are both limited and that *C. albicans* clades do not represent a collection of cryptic species. The study of haplotypes at heterozygous loci performed on our dataset indicates that loss of heterozygosity events due to mitotic recombination is moderately common in natural populations of *C. albicans*. The maintenance of substantial heterozygosity despite relatively frequent loss of heterozygosity could result from a selective advantage conferred by heterozygosity.

Keywords

Multi-Locus Sequence Typing; Loss of heterozygosity; mating; Ca3; clade; recombination

*Corresponding author: Christophe d'Enfert, Unité Biologie et Pathogénicité Fongiques, Institut Pasteur, 25, rue du Docteur Roux, 75724 Paris Cedex 15, France, Tel: +33 (0)1 40 61 32 57, Fax: +33 (0)1 45 68 89 38, E-mail: denfert@pasteur.fr.
[§]MEB and CP contributed equally to this work

Introduction

In many pathogenic microorganisms, genetic exchanges are limited. This may reflect a common strategy facilitating the maintenance of adapted clones while retaining the ability to generate and select new gene combinations in response to changes in selective pressure (Heitman, 2006). The yeast *Candida albicans*, the most prevalent fungal pathogen in humans (Calderone, 2002), appears to exploit such a strategy. *C. albicans* is an obligate diploid with a documented parasexual cycle involving the mating of cells of opposite mating type that have undergone a loss of heterozygosity at the Mating-Type-Like (MTL) locus. This process results in tetraploids (Hull et al., 2000; Magee and Magee, 2000) that can revert to diploidy by random chromosome loss (Bennett and Johnson, 2003). So far, meiosis has not been observed despite the presence in the genome of numerous orthologues of the *Saccharomyces cerevisiae* genes necessary for mating, meiosis and sporulation (Tzung et al., 2001). The shuffling of genetic material and the recombination events that would result from mating are possible sources of diversity in *C. albicans* that could contribute to its success as a pathogen. In this regard, evaluating the contribution of mating to genome diversity within the *C. albicans* species is of significant importance.

Genetic studies of the population structure of *C. albicans* using diverse approaches have all reached a similar general conclusion, namely that the propagation of this pathogen is largely clonal (Pujol et al., 1993; Boerlin et al., 1996; Gräser et al., 1996; Pujol et al., 1997; Cowen et al., 1999; Forche et al., 1999; Lott et al., 1999; Taylor et al., 1999; Xu et al., 1999a; Xu et al., 1999b; Anderson et al., 2001; Fundyga et al., 2002; Tavanti et al., 2004). The results of these studies suggest that mating may be a rare event. Observed recombination might reflect events of mitotic recombination between homologous chromosomes or parasexual exchange rather than true sexual reproduction (Odds et al., 2007). There is, however, an alternative hypothesis. Mating may occur frequently, but may be limited to strains that are genetically related and thus would not lead to measurable recombination (Soll and Pujol, 2003). In the majority of studies on the population structure of *C. albicans*, the samples complexities were largely neglected. *C. albicans* was treated as a homogeneous population. In recent years, this assumption has been shown to be an oversimplification. *C. albicans* can be subdivided into distinct clonal lineages. The existence of clades was initially demonstrated by a comparison of three independent genotyping methods (Pujol et al., 1997), Ca3 fingerprinting, random amplification of polymorphic DNA (RAPD), and multilocus enzyme electrophoresis (MLEE). Each of these methods was shown to identify the same three groups of strains, subsequently termed clades I, II, and III. Ca3 fingerprinting was then used to separate worldwide collections of *C. albicans* into the five major genetic groups or clades, I, II, III, SA, and E (Pujol et al., 1997; Blignaut et al., 2002; Pujol et al., 2002; Soll and Pujol, 2003). More recently, MultiLocus Sequence Typing (MLST) confirmed the existence of the same major clades (Robles et al., 2004; Tavanti et al., 2005; Odds et al., 2007) and indicated that *C. albicans* could be subdivided into additional minor clades (Tavanti et al., 2005; Odds et al., 2007).

Because population genetic analyses were performed without considering subgroups, the possibility existed that while recombination was rare in *C. albicans* between isolates from different clades, it could be common within each clade (Pujol et al., 2005). The clades appear to be reproductively isolated and may, therefore, represent cryptic species. In a cryptic species model, genetic exchange would be common within, but rare between, reproductively isolated populations (Maynard Smith et al., 1993). We, therefore, performed a study of the population structure of *C. albicans* in which clades were assessed separately.

Among the DNA typing methods used to assess the population structure of *C. albicans*, MLST provides a very efficient approach (Bougnoux et al., 2002; Bougnoux et al., 2003; Tavanti et al., 2003; Bougnoux et al., 2004; Robles et al., 2004; Tavanti et al., 2004; Tavanti et al.,

2005; Odds et al., 2007). It is based on the nucleotide sequences of internal regions of 6 or 7 housekeeping genes. The combination of the genotypes obtained at each locus characterizes the Diploid Sequence Type (DST) for each strain. MLST provides reproducible data that are easily obtained, stored and shared. Recent reports (Robles et al., 2004; Tavanti et al., 2005; Odds et al., 2007) have demonstrated that its effectiveness in discriminating between clades is similar to that of Ca3 fingerprinting. Using MLST, we have tested the structure within clades and have identified haplotypes (haploid genotypes). Hardy-Weinberg analysis suggests that genetic exchange is limited not only between clades, but also within clades, and hence, that clades do not represent cryptic species. Haplotype analysis suggests that loss of heterozygoty due to mitotic recombination is relatively common in *C. albicans*, but is not sufficient to upset a general pattern of excesses of heterozygotes.

Materials and Methods

C. albicans isolates

A collection of 203 *C. albicans* isolates was studied (Supplemental Table 1) that included (*i*) a set of 77 unrelated strains previously characterized using DNA fingerprinting with the Ca3 probe (Pujol et al., 1997; Blignaut et al., 2002; Pujol et al., 2002), and (*ii*) a set of 125 unrelated strains collected in Europe between 1998 and 2002 comprising 117 isolates from humans and 8 from birds (Starlings). The latter isolates were collected from seven sites in Europe. The reference strain SC5314 (Fonzi and Irwin, 1993) was also included. The reference collection of 77 isolates (Pujol et al., 1997; Blignaut et al., 2002; Pujol et al., 2002) contained 24 strains from clade I, 14 from clade II, 15 from clade III, 10 from clade SA, 6 from clade E and 8 outliers that did not cluster in any of the five identified major clades. The isolates were stored at − 80°C in Cryo-bille tubes (Laboratoire AES, Combourg, France) until they were grown for genomic DNA extraction as described (Bougnoux et al., 2002).

MLST Typing

MLST was performed as described by Bougnoux *et al.* (2002). Briefly, this involved amplifying and sequencing portions of the coding regions of six housekeeping genes (*ACC1*, 407bp; *VPS13*, 403bp; *GLN4*, 404bp; *ADP1*, 443bp; *RPN2*, 306bp; *SYA1*, 391bp). Because *C. albicans* is diploid, the IUPAC nomenclature one-letter code for nucleotides was used to distinguish homozygous and heterozygous polymorphic sites. New and already described genotypes (ST for sequence type) at each locus were assigned numbers in accordance with the nomenclature used in the *C. albicans* MLST database (http://calbicans.mlst.net, Bougnoux et al., 2004) that follows the one used in Bougnoux *et al.* (2002). The combination of genotypes obtained at the 6 loci characterized the diploid sequence type (DST) for each strain (Supplemental Table 1). Alternatively, the reference collection of 77 strains was typed according to Bougnoux *et al*. (2003) through amplification of seven loci (*AAT1a*, 373 bp; *ACC1*, 407bp; *VPS13*, 403bp; *ADP1*, 443bp; *SYA1*, 391bp; *MPIb*, 375 bp; *ZWF1b*, 491 bp). These data are available in Supplemental Table 2. Additional MLST data obtained by Odds *et al*. (2007) were retrieved from http://test1.mlst.net. Haplotypes were inferred from MLST heterozygous genotypes through subcloning of the PCR products into pCR®2.1-TOPO (TOPO TA Cloning® Invitrogen, Life Technologies, France), amplification of the cloned fragments from 3–8 Lac^- colonies and direct sequencing of the resulting PCR products. Alternatively, haplotypes were inferred with PHASE 2.1 (Stephens and Scheet, 2005).

Data Analysis

MLST identifies polymorphic sites within each of the six sequenced loci. Importantly, we observed that each polymorphic site is distinguished by only three alternative configurations: two that are homozygous (*e.g.*, A/A or G/G) and one that is heterozygous (*e.g.*, A/G; see Supplemental Table 2). Hence, for each isolate, the nucleotides corresponding to the

polymorphic sites found for the 6 loci were translated into a code where 0 stands for one homozygous configuration, 1 stands for the other homozygous configuration and 0.5 stands for the heterozygous configuration. These represent gene frequencies for one of the two alleles and were used to generate a dendrogram from matrixes of pairwise Nei's genetic distances using the Unweighted Pair Group Method with Arithmetic averages (UPGMA) computed by the PHYLIP package, version 3.62 (http://evolution.genetics.washington.edu/phylip.html). Bootstrapping with 1000 randomizations was used to assess the significance of the nodes. Linkage disequilibrium was assessed using the Index of Association (Ia, Maynard Smith et al., 1993) as calculated by the Multilocus 1.3 software package, available at http://www.agapow.net/software/multilocus/(Agapow and Burt, 2001). These tests are independent of the assumption of ploidy level. The levels of significance for nonrandom association between loci were computed under the null hypothesis of a freely recombining population (panmixia). The approach used to assess the levels of significance of linkage disequilibrium was similar to that previously used on populations of pathogenic protozoa and fungi (Tibayrenc et al., 1990; Tibayrenc et al., 1991; Pujol et al., 1993; Burt et al., 1996). Frequencies for heterozygous genotypes expected at Hardy-Weinberg equilibrium were calculated at http://www.changbioscience.com/genetics.hardy.html using values available for homozygous genotypes.

Chromosomal Location of the Loci

Sequenced gene fragments were localized to contigs of the most recent assembly of the *C. albicans* genome sequence from the Stanford Genome Technology Center (Assembly 19, Jones et al., 2004) by performing a blast search at http://genolist.pasteur.fr/CandidaDB (d'Enfert et al., 2005). The genomic location of the loci and their relative orientation in function of the putative centromeres were deduced from the *C. albicans* genomic map browser and contig assembly website (http://candida.bri.nrc.ca/candida/alignments/index.cfm). Locations of the putative centromeres were deduced from Sanyal *et al.* (2004).

Results

MLST analysis of 203 C. albicans strains and characterization of the clades

Analysis of a collection of 203 isolates by MLST over six gene sequences identified 68 polymorphic nucleotide positions that defined 157 diploid sequence types (Supplemental Table 1). One DST was shared by 16 isolates, one by 5 isolates, three by 4 isolates each, three by 3 isolates each, and twelve by 2 isolates each. Each of the remaining 137 DSTs was represented by a single isolate. Among the 203 isolates analyzed, 77 were previously analyzed with the Ca3 fingerprinting method and were representative of the five major clades previously described (Pujol et al., 1997; Blignaut et al., 2002; Pujol et al., 2002). A UPGMA dendrogram was generated (Fig. 1), and the latter set of isolates was used as a reference to identify genetic groups within the larger collection. Due to the overall very close relatedness of *C. albicans* isolates and the possible occurrence of mating in natural populations, the UPGMA analysis was not an attempt to resolve an intraspecies phylogeny, but was rather used as a pragmatic approach to cluster strains and identify clades. In this context, bootstrap values of more than 50% are presented in Figure 1 to indicate the most clearly defined groups of strains but should not be considered to imply genuine phylogenetic support. Three of the 77 strains showed a discrepancy in clustering between the two methods. The II-62 strain, which represents the reference strain WO-1, was outside of clade II and clustered with a single strain. Strains III-27 and III-43 clustered separately from clade III. Nevertheless, strains displaying distinct clustering with the two methods were rare and the clustering of the reference collection closely mirrored the clades previously identified by Ca3 analysis (Fig. 1). Around 96% of the strains identified in one of the clades defined by Ca3 fingerprinting clustered in its respective clade by MLST. With the exception of group III, the clusters identifying each of the major genetic

groups were supported by bootstrap values of 50% or more. In contrast, the majority of the remaining nodes were not supported by bootstrap analysis (Fig. 1). This was particularly true for the deep nodes linking the clades together. Bootstrap values of 50% or more were also observed for four additional clusters containing from 5 to 11 strains. We termed these minor clusters, clade IV to VII (Fig. 1). Most of the strains identified as group III by Ca3 analysis also clustered by MLST, suggesting that this cluster represents a genuine discrete genetic group within *C. albicans*, even if bootstrap analysis did not support any clear delineation. For this reason, in subsequent analyses we tested in parallel the collection of 24 group III strains, and a reduced set of 18 group III strains defined at an arbitrary patristic threshold of 0.035 (Fig. 1). This threshold was chosen because it was useful in defining the remaining clades, with the notable exception of clade SA (Fig. 1). The results obtained for the two sets of strains were highly similar and demonstrated the same trends (data not shown). Therefore, we tentatively defined clade III as being represented by the larger collection of 24 strains (Fig. 1).

Nature and distribution of the polymorphisms

We observed only two alleles at each of the 68 polymorphic nucleotide positions. Nine of the polymorphic nucleotide positions were specific to a single strain each (Supplemental Table 3). The majority of the remaining 59 polymorphic nucleotides observed over the 2,354 bp analyzed were synonymous. Twenty-three polymorphisms generated amino acid substitutions (Supplemental Table 3). Among them 21 were conservative replacements involving amino acids with similar functional groups (17 nucleotide positions) or amino acids with similar effect on secondary structure (4 nucleotide positions). Only 2 nucleotide polymorphisms generated amino acids that were notably distinct and more likely to affect gene function. The average number of heterozygous loci per strain was 3.90 ± 1.27 with a range of 1 to 6, and the mean number of heterozygous nucleotide positions per strain was 13.0 ± 6.03 with a range of 1 to 31.

Most of the nucleotide polymorphisms were shared between two or more different clades and/or outlier strains that did not cluster in any of the 9 clades (Supplemental Table 3). Only one polymorphism was restricted to a unique clade. While the *VPS13* sequence was heterozygous A/G at position 33 in all seven strains of clade VII, it was homozygous G/G for the remaining strains (Supplemental Table 3). Instances of homoplasy were not rare and were most notable between many clades and/or outlier strain combinations (Supplemental Table 3). This homoplasy was reflected by the overall poor bootstrap support observed (Fig. 1).

The chromosomal location of the loci used in this study could affect some of our analyses if loci were too close to each other or in proximity of the centromeres. To test for physical linkage between loci, we verified their chromosomal location. Chromosomal location of the 6 loci were as follows: *ACC1* and *ADP1*, chromosome R; *RPN2*, chromosome 1; *VPS13*, chromosome 4; *SYA1*, chromosome 6; *GLN4*, chromosome 7. *ACC1* and *ADP1* were approximately 1,000 kb apart. *SYA1* was located at approximately 180 kb from the chromosome 6 centromere. All other loci were located between 300 kb and 1,500 kb from a centromere.

Correlation between the Ca3 and MLST Methods and Linkage Disequilibrium

The Mantel nonparametric test was used to compare the nonrandom association between the similarity coefficient matrices obtained for the 77 isolates fingerprinted with the Ca3 method and the Nei's genetic distance matrix obtained for MLST. The Mantel test was highly significant ($P < 10^{-4}$), supporting a strong correlation between the two methods. Correlation between two sets of independent genetic markers represents robust evidence suggesting linkage disequilibrium and a clonal population structure (Pujol et al., 1997). We further assessed linkage disequilibria by using statistics implemented in the Multilocus 1.3 software (Agapow and Burt, 2001). Each of these statistics tested the null hypothesis of a freely recombining

(panmictic) population (Supplemental Table 4). Highly significant linkage disequilibria ($p < 10^{-5}$) were found for both the total collection of 203 isolates and the reduced collection of 157 isolates that did not contain repeated genotypes (Supplemental Table 4). These results indicate that the overall population structure of *C. albicans* is clonal.

Haplotype analysis

While recombination is a source of genetic variation that is predicted to play an important role in *C. albicans*, this term has been used to describe different mechanisms including genetic exchange and mitotic recombination. In order to estimate the contribution of mitotic recombination to the genetic diversity of *C. albicans* we performed a haplotype analysis within each of the genetic groups previously described. The study was performed on 150 strains from the six clades represented by more than 10 isolates. These strains included 52 clade I isolates, 16 clade II isolates, 24 clade III isolates, 11 clade IV isolates, 26 clade E isolates, and 21 clade SA isolates. Based on the genotypes obtained at each locus, we deduced the allelic make-up of strains from each genetic group. A representative example is given in Table 1 for the *GLN4* locus analyzed in 52 clade I isolates. Seven genotypes were found in that case. The three genotypes that were the most commonly found ST2, ST9, and ST11 could be interpreted as being the result of the differential combination of two alleles (alleles *1* and *2* in Table 1). While isolates presenting the ST2 and ST9 genotypes were homozygous for alleles *1* and *2*, respectively, isolates with the ST11 genotype were interpreted to be *1/2* heterozygotes. The four rare genotypes represented a minority of group I isolates (8%) and were interpreted to be recombinant genotypes derived from ST11 *1/2* heterozygous strains that underwent a loss of heterozygosity through recombination between alleles *1* and *2* (Table 1). As a result of these recombinations, “hybrid” alleles that were part allele *1* and part allele *2* emerged. A conservative interpretation of these data suggests that the recombinant genotypes reflect the combination of one “hybrid” allele and one “original” allele in cases of partial loss of heterozygosity over the locus. This allelic interpretation was verified by cloning and sequencing individual *GLN4* alleles from 28 group I strains that were heterozygous at this locus. The number of clones analyzed per strain (three to eight) was too small to always recover both alleles, but in cases where only one allele was sequenced it corresponded to one of the deduced alleles. While in 10 of the 25 *GLN4*-ST11 strains analyzed the two expected alleles (alleles *1* and *2* in Table 1) were found, only one of the two alleles (*1* or *2*) was retrieved in the remaining 15 strains. The two expected alleles (Table 1) were recovered from the strains presenting the *GLN4*-ST14, ST16, and ST21 recombinant genotypes. These results confirmed our allelic interpretation. The same approach that was used to deduce the clade I *GLN4* haplotypes was also applied to determine haplotypes within each clade and for all loci. Allelic interpretations were again confirmed on select strains, for all loci, by cloning and sequencing individual alleles. Haplotype results are summarized in Supplemental Table 5. Alleles were cloned and sequenced for a total of 126 gene sequences. In all cases, allele sequences were in agreement with our allelic interpretations. Moreover, our allelic interpretations were in general agreement with haplotypes predicted using the PHASE 2.1 software (Stephens and Scheet, 2005). Out of 56 different single locus genotypes with more than one heterozygous nucleotide position, 50 were attributed identical haplotypes by both approaches. Only in 6 (11%) instances (data not shown) did the predictions differ. In these cases, PHASE 2.1 prediction confidence was generally low and the predicted haplotypes differed from the sequencing data obtained for the corresponding alleles except in one case where the allelic combination had not been validated by sequencing (data not shown). A total of 37 recombinant genotypes composed of variant alleles generated by mitotic recombination were observed on 150 strains analyzed for the six MLST loci (Table 2). We interpreted this data as evidence of a relatively frequent loss of heterozygosity over the different loci.

Excesses of heterozygotes are common within C. albicans clades

Despite the number of studies dealing with the population structure of *C. albicans*, no study has attempted to test the population structure within clades. We, therefore, used the allelic frequencies obtained at each polymorphic nucleotide position to assess non-random segregation within clades. Hardy-Weinberg equilibrium was tested to ascertain whether the observed genotypic frequencies were consistent with those expected from random mating (Table 3 and Supplemental Table 6). Within each clade, genotypic frequencies were grossly inconsistent with the expected Hardy-Weinberg frequencies at multiple nucleotide positions located on most of the loci. When strains from individual clades were analyzed, in each case of departure from Hardy-Weinberg equilibrium, excess of heterozygotes was observed. In contrast, when isolates from the distinct clades were grouped together, deficit of heterozygotes prevailed (83%; Table 3 and Supplemental Table 6). The collection of isolates analyzed was from diverse origins, including 136 isolates from Europe and 59 from Northern America (Supplemental Table 1). This could have biased our results, because we considered the *C. albicans* collection as representing a single population. The possibility existed that deviations from panmixia observed here might have been the result of grouping together geographically isolated populations. Geographical isolation may be associated with different allelic frequencies in different populations, even if each separate population is panmictic, and may have resulted in apparent departures from Hardy-Weinberg expectations in our collection. If this were the case, however, we should have observed a deficit of heterozygotes (Whalund effect, Wahlund, 1928). The excesses of heterozygotes observed here in each clade and at many loci suggest that our results are not due to a geographical isolation bias. We nevertheless verified this possibility by testing European or North American populations separately whenever possible. The results obtained for geographically isolated populations showed the same trend than when clades containing strains from different origins were analyzed (Supplemental Table 7), namely that departure from Hardy-Weinberg expectations within clades were frequent and indicated excesses of heterozygotes in each case. This clearly demonstrated that geographical isolation could not explain our results and that mating within clades is limited.

Our analysis was performed on small sample sizes of between 11 and 52 isolates per clade. These sample sizes may be too small to accurately reflect the general status of *C. albicans*. To ascertain our findings, we analyzed Hardy-Weinberg equilibrium on a larger collection of isolates. Recently Odds et al. (2007) have published a substantial MLST study on more than 1,000 *C. albicans* isolates using a different scheme that analyzes seven loci instead of the six used in our study (Bougnoux et al., 2003). This scheme also differs from the method we used in that only four of the loci are common to both scheme (*ACC1*, *VPS13*, *ADP1*, and *SYA1*). In their work, Odds et al. (2007) had also shown that the MLST approach they used identified clades that were in agreement with the clades described by Ca3 fingerprinting. We first needed to identify which of their clades corresponded to our clades. To realize that, we typed the reference collection of 77 isolates initially fingerprinted with Ca3, with the consensus MLST scheme. The DSTs obtained were then compared to the Odds et al. (2007) dataset to identify clade membership. The results are presented in Supplemental Table 2. All of the 24 isolates that were included in clade I clustered in the clade 1 described by Odds et al. (2007). Twelve of the 13 clade II isolates clustered in clade 2. The remaining clade II isolate did not cluster in any of the clades described by Odds et al. (2007). Eleven of the 13 clade III isolates clustered in clade 3, one clustered in clade 1, and one did not cluster in any of the clades described by Odds et al. (2007). The 10 clade SA isolates clustered in clade 4, the 6 clade E isolates clustered in clade 11, and the two clade IV isolates clustered in clade 9. This results suggested that clades I, II, III, SA, E, and IV corresponded to clades 1, 2, 3, 4, 11, and 9 from Odds et al. (2007), respectively. Overall, close to 95% of the strains identified as being members of one clade by one MLST approach were found to cluster in its corresponding clade by the second approach,

indicating a reasonable correspondence between the two approaches. The Mantel test supported a strong correlation between the three methods ($P < 10^{-4}$ in each case). Because the correspondence between clades IV and 9 was based on only two strains, clade 9 was not included in further analyses. Data for the 1,044 strains from clades 1, 2, 3, 4, and 11 described by Odds et al. (2007) were retrieved from http://test1.mlst.net for further analysis. Analysis of linkage disequilibrium performed on this sample (Supplemental Table 4) indicated again highly significant linkage disequilibria ($P < 10^{-5}$) over the entire sample. Non-random segregation within clades and over the entire sample was performed by testing Hardy-Weinberg equilibrium for each polymorphic nucleotide position. The results presented in Table 4 and Supplemental Table 8 again indicated that while excess heterozygotes prevailed within clades (96%), deficits of heterozygotes were preponderant over the entire sample (83%). The proportion of polymorphic nucleotide sites with significant excess heterozygotes found in each clade was analyzed for this large dataset (Table 4). The number of polymorphic nucleotide positions was determined for each individual clade. Sites were considered to be significantly polymorphic whenever the major allele frequency was lower than 0.95. Sites with a major allele frequency of more than 0.95 were virtually monomorphic and not used in this analysis. Between 50.0% and 85.2% of the polymorphic sites had significant excess heterozygotes in the five clades analyzed (Table 4). The mean number of sites with excess heterozygotes was 66.2% ± 12.9. In addition, for the four loci common to the two MLST schemes, among the 64 polymorphic nucleotide positions that showed significant excess or deficit of heterozygotes in the collection of 150 isolates (Supplemental Table 6), 59 showed similar deviation from Hardy-Weinberg in the collection of 1,044 isolates, while the remaining 5 did not show any deviation (Supplemental Table 8). This dataset differed markedly from our original data. The collection of isolates was significantly larger, was analyzed by using a different MLST scheme, and the clades were identified by Odds et al. (2007) using their own criteria. Nevertheless, the results obtained were remarkably similar and demonstrated the same trend than in our initial analysis. This expanded analysis emphasizes and generalizes our finding and suggests that our results are not likely the consequence of artifacts due to sampling, molecular markers, or strain clustering.

Discussion

The population structure of *C. albicans* has been the subject of numerous studies using a variety of, sometimes unrelated, approaches and analyses. Each of these studies have reached the same general conclusion, namely that the population structure of *C. albicans* was largely clonal (Pujol et al., 1993; Gräser et al., 1996; Anderson et al., 2001; Tavanti et al., 2004; Pujol et al., 2005) and that mating, if it occurred, was rare or essentially limited to related strains and thus difficult to document (Soll and Pujol, 2003; Pujol et al., 2005). The demonstration of congruence among three independent fingerprinting methods, Ca3 fingerprinting, RAPD, and MLEE, represented a particularly compelling evidence in favor of a clonal population structure in *C. albican*s (Pujol et al., 1997). In this study, we have investigated the population structure of a collection of 203 *C. albicans* isolates using MLST and expanded this analysis to an additional collection of 1,044 isolates. Importantly, 77 of these isolates had already been typed using Ca3 fingerprinting (Pujol et al., 1997; Blignaut et al., 2002; Pujol et al., 2002) and our results demonstrate a good agreement between two MLST typing schemes and Ca3 typing, with the same major clades identified by the three approaches. This congruence between independent genetic markers provides additional robust evidence for linkage disequilibrium and clonality in *C. albicans* (Pujol et al., 1997).

The recent characterization of discrete genetic groups in *C. albicans* has been of interest not only to the global epidemiology of this species (Pujol et al., 1997; Blignaut et al., 2002; Pujol et al., 2002; Soll and Pujol, 2003), but also for identifying clade-specific phenotypic traits, most notably antifungal resistance (Dodgson et al., 2004; Pujol et al., 2004). Five major genetic

groups were originally identified by using independent fingerprinting methods, which were subsequently shown to exhibit differences in geographical distribution (Blignaut et al., 2002; Pujol et al., 2002) and more recent studies revealed additional minor groups (Tavanti et al., 2005; Odds et al., 2007), similar to what we have observed here. It was hypothesized that these clonal lineages paralleled the evolution of distinct phenotypic traits relevant to pathogenesis or drug resistance (Soll and Pujol, 2003). Clade-specific resistance to flucytosine (Dodgson et al., 2004; Pujol et al., 2004) and Amphotericin B (Blignaut et al., 2005), and differences in adhesin (ALSs) structure that could impact function were indeed demonstrated (reviewed in Pujol et al., 2005). In the present study, we have investigated whether clades could represent cryptic species within which mating occurred while the overall *C. albicans* species would still present a clonal population structure. This had not been tested so far. To investigate this hypothesis, a large sample that included critical numbers of strains within each clade was essential. In each of the clades analyzed with the 6 loci MLST scheme, polymorphic nucleotide sites located on 3 or 4 of the 6 loci were found to be in Hardy-Weinberg disequilibrium. Each of these loci proved to have an excess of heterozygotes. Similarly, in the analysis performed on the larger collection of strains typed with the 7 loci MLST scheme, polymorphic sites located on 5 to 7 of the 7 loci were in Hardy-Weinberg disequilibrium within clades. The majority of these polymorphisms showed excess heterozygotes (96 out of 100) as well, and close to 66% of all polymorphic positions in the individual clades were in significant excess of heterozygotes over the five clades. This is different from previous reports in which whole collections have been analyzed (Pujol et al., 1993; Boerlin et al., 1996; Gräser et al., 1996; Cowen et al., 1999; Forche et al., 1999; Lott et al., 1999; Xu et al., 1999a; Xu et al., 1999b; Fundyga et al., 2002; Tavanti et al., 2004). In these studies, deficits of heterozygotes were revealed more frequently than excesses. In our hands, Hardy-Weinberg analyses performed on the overall collection rather than within clades also demonstrated a prevalence of deficits of heterozygotes. These were to some extent due to a significant proportion of positions that were monomorphic in the clades while demonstrating polymorphisms over the entire collection. Ninety one percent and 83% of nucleotide positions in Hardy-Weinberg disequilibrium presented deficits of heterozygotes in the 6 loci and the 7 loci MLST schemes, respectively. This indicates that the seemingly difference in trends between this study and previous studies was not due to an intrinsic difference in the genetic markers used, but rather reflects the significance of analyzing clades individually. Our finding that excesses of heterozygotes are frequent among strains from the same clade strongly suggests that mating within clades is limited, indicating in turn that *C. albicans* clades do not represent cryptic species. The results also suggest that selfing (self-fertilization) is not frequent. Otherwise, deficits of heterozygotes would have been observed. Taken together, these results and others (reviewed in Pujol et al., 2005) indicate that genetic exchange is extremely rare in *C. albicans* within clades as well as between clades.

In a recent study, Odds *et al*. (2007) have analyzed *C. albicans* recombination by using founders of the largest eBURST clonal clusters identified from MLST data of 1391 isolates. This analysis indicated high frequency of recombination such that these selected isolates resembled those in a sexually reproducing species. The focus on the founders of clonal clusters and hence the most heterozygous isolates in the strain collection, the small size of the analyzed sample and the lack of clade-specific analyses may explain the apparent discrepancy between the two studies. Moreover, Odds *et al*. (2007) have proposed that their observation may reflect mitotic recombination followed by parasexual exchange rather than mating and true sexual recombination.

Haplotype analyses were performed under the assumption that the DNA sequences containing our six MLST loci were diploid. A recent study by Selmecki et al. (2005) has elegantly demonstrated that the genome of most *C.albicans* strains was diploid. They showed that while the frequency of chromosome pairs presenting aneuploidy was approximately 1% among fluconazole-sensitive strains, in fluconazole-resistant strains it rose to around 10%. They also

found that aneuploidy was most often found on chromosome 5. To limit aneuploidy in our sample, we used a collection of fluconazole-sensitive srains and did not use loci located on chromosome 5. In conclusion, even if close to 1% of our data may be tainted by aneuploidy, such an error margin cannot be put forward to explain our results.

We did analyze 6 loci that represented only a minute fraction of the *C. albicans* genome. Nevertheless, each of the six loci used indicated the same trend. It is highly unlikely that we selected six loci with the same aberrant behavior. This suggests that our findings must reflect a more general trend that should be valid at many more genomic loci. We tested this hypothesis by expanding the analysis of non-random segregation to three new loci (*ACC1*, *MPI*, and *ZWF1*) over 1,044 isolates. For each of these loci we demonstrated that the same pattern of excess heterozygotes held true within clades.

Our results do not negate the possibility that rare instances of mating and genetic exchange may occur. The homoplasy observed for the polymorphisms shared by different clades and/or outlier strains, and the fact that practically all polymorphisms were shared suggest genetic exchange. These homoplasies are unlikely to be due to parallelism because the loci analyzed were not hypervariable, as indicated by the low level of polymorphism and the presence of only two alleles at each polymorphic position. We cannot, however, rule out the possibility that most of these shared polymorphisms were inherited from an ancestral population that frequently mated in the past. Consistent with this possibility, Anderson et al. (2001) have proposed that modern *C. albicans* populations were founded by a few highly heterozygous and sexual ancestors. Contemporary rare events of mating and genetic exchange may also provide an occasional response to changing selective pressures and contribute to generate new clonal populations better fitted to host and environmental niches (Heitman, 2006). Rare mating may explain the few instances of incongruence between the two MLST typing schemes and the Ca3 method (Supplemental Table 2). These could either result from homoplasy associated with a particular genetic marker or may be indicative of rare occurrence of genetic exchange as was suggested by two studies on incongruence between mitochondrial and nuclear markers (Anderson et al., 2001; Wang et al., 2007).

Loss of heterozygosity in *C. albicans* has been documented during the course of infection as well as during commensalism or carriage (Pujol et al., 2005; Bougnoux et al.), and has recently been shown to be involved in the acquisition of resistance to flucytosine (Dodgson et al., 2004) and fluconazole (Coste et al., 2006) as well as the generation of mating-competent strains (Wu et al., 2005; Wu et al., 2007), indicating that it can have a major impact on phenotype. Loss of heterozygosity as evidenced by multiple instances of recombinant genotypes in our study appears to be relatively frequent and to play a role in the evolution of *C. albicans* strains. This is in agreement with two recent MLST-based studies that demonstrated frequent loss of heterozygosity in related isolates from the same individual (Odds et al., 2006) or from different individuals from the same family (Bougnoux et al., 2006). Together these studies indicate that loss of heterozygosity is an ongoing mechanism leading to increased genetic diversity in *C. albicans*. Recurrent loss of heterozygosity will generate new genotypes but should ultimately result in complete homozygosis without a counterbalancing mechanism. Excesses of heterozygotes observed in our study demonstrate that such a mechanism is indeed at play. Three different mechanisms could restore or maintain heterozygosity, namely (*i*) genetic exchange, (*ii*) *de novo* mutations, or (*iii*) selection. Genetic exchange seems too rare to play such a role and should not lead to excesses of heterozygotes. The generation of new mutations is an unlikely candidate for the tested loci because they were not hypervariable. Heterozygosity may be globally maintained in *C. albicans* by selection due to the presence of deleterious alleles or because the maintenance of alternate alleles at some loci confers a selective advantage. This hypothesis does not imply that each nucleotide position in excess of heterozygotes is directly under selective pressure. Indeed, each of the 9 tested loci demonstrated nucleotide positions

in excess of heterozygotes in at least 2 of the clades, and it seems extremely unlikely that we would have randomly selected only loci displaying a heterozygote advantage. In addition, most of the polymorphisms did not lead to an amino acid substitution, and the few substitutions were mostly conservative (ie. similar functional groups or similar predicted effect on secondary structure). Our data rather suggest that frequent heterozygote excesses are indicative of a more global selective pressure to maintain heterozygosity in clades for the benefit of a number, as yet unspecified, of genes which function may be effected by zygosity in the coding or the promoter sequences. This was suggested by two recent studies in *C. albicans* (Lockhart *et al.*, 2005; Wu *et al.*, 2007). Wu et al. (2007) have shown that *MTL*-homozygosis in nature was associated with a loss of heterozygosity over short tracts of DNA encompassing the *MTL* locus. In contrast, strains that spontaneously lost *MTL*-heterozygosity *in vitro* underwent homozygosis over large regions of DNA, mostly through uniparental disomy (loss of one chromosome 5 and duplication of the retained homolog; Wu et al., 2005). *MTL*-homozygous strains were also shown to be more virulent than spontaneous *MTL*-homozygous strains obtained *in vitro* (Wu et al., 2007), suggesting that virulence and fitness in nature may be both closely dependent on the upholding of heterozygosity over at least some loci. This hypothesis is in agreement with our results where despite instances of homozygosis at tested loci, heterozygosities remained frequent and excess of heterozygotes were maintained within clades. Single crossovers (or break-induced replication, Prado et al., 2003) would lead to a loss of heterozygosity at all of the loci distal to the recombination site and centromere, and chromosome loss would lead to the entire loss of heterozygosity over the chromosome. Were strains undergoing single crossovers or chromosome loss maintained in nature, an overall reduction in heterozygosity should result. This was not observed in our study and the most parsimonious hypothesis is that loss of heterozygosity in *C. albicans* may be relatively frequent and those that affect short tracts of DNA are preferentially selected over those affecting larger genomic regions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Thanks are due to Murielle Chauvel, Arnaud Firon, Sophie Goyard, and Maud Lecomte for their insights and contribution to this topic. Work in the laboratory of CE was supported by grants from Pasteur-Génopole Ile-de-France. D. Diogo is the recipient of a PhD fellowship from Institut National de la Recherche Agronomique. This research was supported by NIH grant AI2392 to

Abbreviations

MLST
: MultiLocus Sequence Typing

DST
: Diploid Sequence Type

References

Agapow PM, Burt A. Indices of multilocus linkage disequilibrium. Mol Ecol Notes 2001;1:101–102.

Anderson JB, Wickens C, Khan M, Cowen LE, Federspiel N, Jones T, Kohn LM. Infrequent genetic exchange and recombination in the mitochondrial genome of *Candida albicans*. J Bacteriol 2001;183:865–872. [PubMed: 11208783]

Bennett RJ, Johnson AD. Completion of a parasexual cycle in Candida albicans by induced chromosome loss in tetraploid strains. EMBO J 2003;22:2505–2515. [PubMed: 12743044]

Blignaut E, Molepo J, Pujol C, Soll DR, Pfaller MA. Clade-related amphotericin B resistance among South African Candida albicans isolates. Diagn Microbiol Infect Dis 2005;53:29–31. [PubMed: 16182076]

Blignaut E, Pujol C, Lockhart S, Joly S, Soll DR. Ca3 fingerprinting of Candida albicans isolates from human immunodeficiency virus-positive and healthy individuals reveals a new clade in South Africa. J Clin Microbiol 2002;40:826–836. [PubMed: 11880401]

Boerlin P, Boerlin-Petzold F, Goudet J, Durussel C, Pagani JL, Chave JP, Bille J. Typing *Candida albicans* oral isolates from human immunodeficiency virus-infected patients by multilocus enzyme electrophoresis and DNA fingerprinting. J Clin Microbiol 1996;34:1235–1248. [PubMed: 8727910]

Bougnoux ME, Aanensen DM, Morand S, Theraud M, Spratt BG, d'Enfert C. Multilocus sequence typing of Candida albicans: strategies, data exchange and applications. Infect Genet Evol 2004;4:243–252. [PubMed: 15450203]

Bougnoux ME, Diogo D, Francois N, Sendid B, Veirmeire S, Colombel JF, Bouchier C, Van Kruiningen H, d'Enfert C, Poulain D. Multilocus sequence typing reveals intrafamilial transmission and microevolutions of Candida albicans isolates from the human digestive tract. J Clin Microbiol 2006;44:1810–1820. [PubMed: 16672411]

Bougnoux, ME.; Diogo, D.; Pujol, C.; Soll, DR.; d'Enfert, C. Molecular epidemiology and population dynamics in Candida albicans. In: d'Enfert, C.; Hube, B., editors. Candida: comparative and fucntional genomics. Caister Academic Press; Wymondham, United Kingdom: 2007. p. 51-70.

Bougnoux ME, Morand S, d'Enfert C. Usefulness of multilocus sequence typing for characterization of clinical isolates of Candida albicans. J Clin Microbiol 2002;40:1290–1297. [PubMed: 11923347]

Bougnoux ME, Tavanti A, Bouchier C, Gow NA, Magnier A, Davidson AD, Maiden MC, D'Enfert C, Odds FC. Collaborative consensus for optimized multilocus sequence typing of Candida albicans. J Clin Microbiol 2003;41:5265–5266. [PubMed: 14605179]

Burt A, Carter DA, Koenig GL, White TJ, Taylor JW. Molecular markers reveal cryptic sex in the human pathogen Coccidioides immitis. Proc Natl Acad Sci U S A 1996;93:770–773. [PubMed: 8570632]

Calderone, R. Candida and candidiasis. ASM Press; Washington DC: 2002.

Coste A, Turner V, Ischer F, Morschhauser J, Forche A, Selmecki A, Berman J, Bille J, Sanglard D. A mutation in Tac1p, a transcription factor regulating CDR1 and CDR2, is coupled with loss of heterozygosity at chromosome 5 to mediate antifungal resistance in Candida albicans. Genetics 2006;172:2139–2156. [PubMed: 16452151]

Cowen LE, Sirjusingh C, Summerbell RC, Walmsley S, Richardson S, Kohn LM, Anderson JB. Multilocus genotypes and DNA fingerprints Do not predict variation in azole resistance among clinical isolates of Candida albicans. Antimicrob Agents Chemother 1999;43:2930–2938. [PubMed: 10582885]

d'Enfert C, Goyard S, Rodriguez-Arnaveilhe S, Frangeul L, Jones L, Tekaia F, Bader O, Albrecht A, Castillo L, Dominguez A, Ernst J, Fradin C, Gaillardin C, Garcia-Sanchez S, de Groot P, Hube B, Klis FM, Krishnamurthy S, Kunze D, Lopez M-C, Mavor A, Martin N, Moszer I, Onésime D, Perez Martin J, Sentandreu R, Valentin E, Brown AJP. CandidaDB: a genome database for Candida albicans pathogenomics. Nucl Acids Res 2005;33(Database Issue):D353–357. [PubMed: 15608215]

Dodgson AR, Dodgson KJ, Pujol C, Pfaller MA, Soll DR. Clade-specific flucytosine resistance is due to a single nucleotide change in the FUR1 gene of Candida albicans. Antimicrob Agents Chemother 2004;48:2223–2227. [PubMed: 15155225]

Fonzi WA, Irwin MY. Isogenic strain construction and gene mapping in *Candida albicans*. Genetics 1993;134:717–728. [PubMed: 8349105]

Forche A, Schonian G, Graser Y, Vilgalys R, Mitchell TG. Genetic structure of typical and atypical populations of Candida albicans from Africa. Fungal Genet Biol 1999;28:107–125. [PubMed: 10587473]

Fundyga RE, Lott TJ, Arnold J. Population structure of Candida albicans, a member of the human flora, as determined by microsatellite loci. Infect Genet Evol 2002;2:57–68. [PubMed: 12798001]

Gräser Y, Volovsek M, Arrington J, Schönian G, Presber W, Mitchell TG, Vilgalys R. Molecular markers reveal that population structure of the human pathogen *Candida albicans* exhibits both clonality and recombinaison. Proc Natl Acad Sci USA 1996;93:12473–12477. [PubMed: 8901606]

Heitman J. Sexual reproduction and the evolution of microbial pathogens. Curr Biol 2006;16:R711–725. [PubMed: 16950098]

Hull CM, Raisner RM, Johnson AD. Evidence for mating of the "asexual" yeast Candida albicans in a mammalian host. Science 2000;289:307–310. [PubMed: 10894780]

Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, Newport G, Thorstenson YR, Agabian N, Magee PT, Davis RW, Scherer S. The diploid genome sequence of Candida albicans. Proc Natl Acad Sci U S A 2004;101:7329–7334. [PubMed: 15123810]

Lockhart SR, Wu W, Radke JB, Zhao R, Soll DR. Increased virulence and competitive advantage of a/alpha over a/a or alpha/alpha offspring conserves the mating system of Candida albicans. Genetics 2005;169:1883–1890. [PubMed: 15695357]

Lott TJ, Holloway BP, Logan DA, Fundyga R, Arnold J. Towards understanding the evolution of the human commensal yeast Candida albicans. Microbiology 1999;145:1137–1143. [PubMed: 10376829]

Magee BB, Magee PT. Induction of mating in Candida albicans by construction of MTLa and MTLalpha strains. Science 2000;289:310–313. [PubMed: 10894781]

Maynard Smith J, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria ? Proc Natl Acad Sci U S A 1993;90:4384–4388. [PubMed: 8506277]

Odds FC, Bougnoux ME, Shaw DJ, Bain JM, Davidson AD, Diogo D, Jacobsen MD, Lecomte M, Li SY, Tavanti A, Maiden MC, Gow NA, d'Enfert C. Molecular phylogenetics of Candida albicans. Eukaryot Cell 2007;6:1041–1052. [PubMed: 17416899]

Odds FC, Davidson AD, Jacobsen MD, Tavanti A, Whyte JA, Kibbler CC, Ellis DH, Maiden MC, Shaw DJ, Gow NA. Candida albicans strain maintenance, replacement, and microvariation demonstrated by multilocus sequence typing. J Clin Microbiol 2006;44:3647–3658. [PubMed: 17021093]

Prado F, Cortes–Ledesma F, Huertas P, Aguilera A. Mitotic recombination in Saccharomyces cerevisiae. Curr Genet 2003;42:185–198. [PubMed: 12589470]

Pujol, C.; Dodgson, AR.; Soll, DR. Population genetics of ascomycetes pathogenic to humans and animals. In: Xu, J., editor. Evolutionary Genetics of Fungi. Horizon Scientific Press; Norfolk, UK: 2005. p. 149-188.

Pujol C, Joly S, Lockhart SR, Noel S, Tibayrenc M, Soll DR. Parity among the randomly amplified polymorphic DNA method, multilocus enzyme electrophoresis, and southern blot hybridation with the moderately repetitive DNA probe Ca3 for fingerprinting *Candida albicans*. J Clin Microbiol 1997;35:2348–2358. [PubMed: 9276415]

Pujol C, Pfaller M, Soll DR. Ca3 fingerprinting of Candida albicans bloodstream isolates from the United States, Canada, South America, and Europe reveals a European clade. J Clin Microbiol 2002;40:2729–2740. [PubMed: 12149321]

Pujol C, Pfaller MA, Soll DR. Flucytosine resistance is restricted to a single genetic clade of Candida albicans. Antimicrob Agents Chemother 2004;48:262–266. [PubMed: 14693548]

Pujol C, Reynes J, Renaud F, Raymond M, Tibayrenc M, Ayala FJ, Janbon F, Mallie M, Bastide JM. The yeast Candida albicans has a clonal mode of reproduction in a population of infected human immunodeficiency virus-positive patients. Proc Natl Acad Sci U S A 1993;90:9456–9459. [PubMed: 8415722]

Robles JC, Koreen L, Park S, Perlin DS. Multilocus sequence typing is a reliable alternative method to DNA fingerprinting for discriminating among strains of Candida albicans. J Clin Microbiol 2004;42:2480–2488. [PubMed: 15184424]

Sanyal P, Baum M, Carbon J. Centromeric DNA sequences in the pathogenic yeast Candida albicans are all different and unique. Proc Natl Acad Sci U S A 2004;101:11374–11379. [PubMed: 15272074]

Selmecki A, Bergmann S, Berman J. Comparative genome hybridization reveals widespread aneuploidy in Candida albicans laboratory strains. Mol Microbiol 2005;55:1553–1565. [PubMed: 15720560]

Soll DR, Pujol C. Candida albicans clades. FEMS Immunol Med Microbiol 2003;39:1–7. [PubMed: 14556989]

Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet 2005;76:449–462. [PubMed: 15700229]

Tavanti A, Davidson AD, Fordyce MJ, Gow NA, Maiden MC, Odds FC. Population structure and properties of Candida albicans, as determined by multilocus sequence typing. J Clin Microbiol 2005;43:5601–5613. [PubMed: 16272493]

Tavanti A, Gow NA, Maiden MC, Odds FC, Shaw DJ. Genetic evidence for recombination in Candida albicans based on haplotype analysis. Fungal Genet Biol 2004;41:553–562. [PubMed: 15050544]

Tavanti A, Gow NA, Senesi S, Maiden MC, Odds FC. Optimization and validation of multilocus sequence typing for Candida albicans. J Clin Microbiol 2003;41:3765–3776. [PubMed: 12904388]

Taylor JW, Geisser DM, Burt A, Koufopanou V. The evolutionary biology and population genetic underlying fungal strain typing. Clin Microbiol Rev 1999;12:126–146. [PubMed: 9880478]

Tibayrenc M, Kjellberg F, Arnaud J, Oury B, Breniere SF, Darde ML, Ayala FJ. Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. Proc Natl Acad Sci U S A 1991;88:5129–5133. [PubMed: 1675793]

Tibayrenc M, Kjellberg F, Ayala FJ. A clonal theory of parasitic protozoa: the population structures of Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas, and Trypanosoma and their medical and taxonomical consequences. Proc Natl Acad Sci U S A 1990;87:2414–2418. [PubMed: 2320563]

Tzung KW, Williams RM, Scherer S, Federspiel N, Jones T, Hansen N, Bivolarevic V, Huizar L, Komp C, Surzycki R, Tamse R, Davis RW, Agabian N. Genomic evidence for a complete sexual cycle in Candida albicans. Proc Natl Acad Sci U S A 2001;98:3249–3253. [PubMed: 11248064]

Wahlund S. Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas 1928;11:65–106.

Wang H, Guo H, Sun S, Xu J. Abundant sequence variation around the mitochondrial origin of replication in the human opportunistic yeast pathogen *Candida albicans* from a tropical island in China. Fungal Genet Biol 2007;44:991–1001. [PubMed: 17493848]

Wu W, Lockhart SR, Pujol C, Srikantha T, Soll DR. Heterozygosity of genes on the sex chromosome regulates Candida albicans virulence. Mol Microbiol 2007;64:1587–1604. [PubMed: 17555440]

Wu W, Pujol C, Lockhart SR, Soll DR. Chromosome loss followed by duplication is the major mechanism of spontaneous mating-type locus homozygosis in Candida albicans. Genetics 2005;169:1311–1327. [PubMed: 15654090]

Xu J, Boyd CM, Livingston E, Meyer W, Madden JF, Mitchell TG. Species and genotypic diversities and similarities of pathogenic yeasts colonizing women. J Clin Microbiol 1999a;37:3835–3843. [PubMed: 10565893]

Xu J, Vilgalys R, Mitchell TG. Lack of genetic differentiation between two geographically diverse samples of Candida albicans isolated from patients infected with human immunodeficiency virus. J Bacteriol 1999b;181:1369–1373. [PubMed: 9973371]

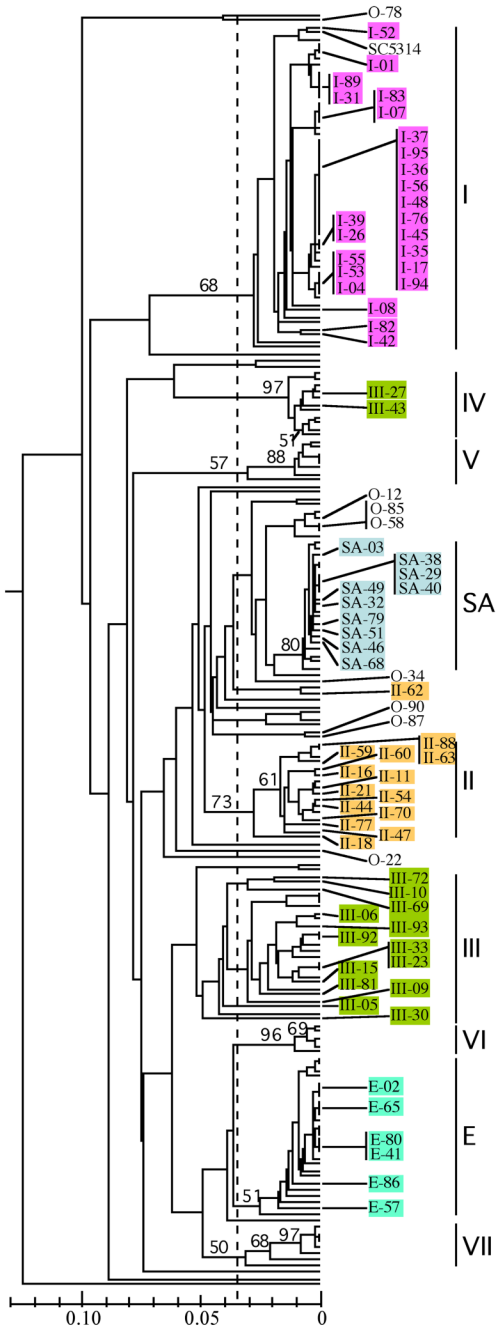

Fig. 1.

Dendrogram based on MLST data obtained for 203 *C. albicans* isolates analyzed at the *ACC1*, *VPS13*, *GLN4*, *ADP1*, *RPN2*, and *SYA1* loci. The dendrogram included the 125 European isolates, the strain SC5314, and the reference collection of 77 isolates previously analyzed by Ca3 fingerprinting. The position of the 77 reference isolates and SC5314 are indicated. The isolates belonging to the five major clades identified by Ca3 fingerprinting are indicated by a prefix; I-n, isolates identified in clade I by Ca3 fingerprinting (magenta background); II-n, clade II isolates (orange background); III-n, clade III isolates (green background); SA-n, clade SA isolates (blue background); E-n, clade E isolates (turquoise background); O-n, outliers that did not cluster in any of the five main genetic groups with the

Ca3 fingerprinting method. The dashed line corresponds to the patristic distance threshold of 0.035. Bootstrap values (1,000 reiterations) of ≥ 50% are shown for nodes clustering more than 3 strains. Four new minor clades, IV to VII, are indicated. The minor clades V, VI and VII contained less than 10 isolates each and were not further studied in this work.

Table 1

Allelic Interpretation of the *GLN4* Genotypes Obtained in 52 Clade I Strains

***GLN4* sequence type**	**Polymorphic nucleotide sites[a]**	**No. of clade I strains (percent)**		**Allelic interpretation**	**Allelic sequences[b]**
ST2	AAGCA	6 (12%)		*1/1*	*1*: AAGCA
ST9	TGAAC	12 (24%)	92%	*2/2*	*2*: TGAAC
ST11	WRRMM	30 (56%)		*1/2*	
ST16	ARRMM	1 (2%)		*1/3*	*3*: AGAAC
ST21	WRGCA	1 (2%)		*1/4*	*4*: TGGCA
ST14	TRRMM	1 (2%)		*2/5*	*5*: TAGCA
ST26	TAGCA	1 (2%)		*5/5*	

[a] The five sites that are polymorphic in *GLN4* for group I isolates. The sites correspond to positions 25, 127, 387, 391, and 393 of the sequenced fragment, respectively. Underlined positions indicate homozygous nucleotides that are interpreted as the result of a loss of heterozygosity due to recombination between alleles *1* and *2*. Locus *GLN4* is 5′ proximal to the centromere.

[b] Underlined nucleotides indicate recombinant positions in “hybrid” alleles.

Table 2

Variant Genotypes Indicative of Recombination.

Genes	Total No. recombinant genotypes for each clade*						
	I 52	**II 16**	**III 24**	**IV 11**	**E 26**	**SA 21**	**Total 150**
ACC1	0	0	1	0	0	0	1
VPS13	4	2	4	3	1	0	14
GLN4	4	0	0	2	0	0	6
ADP1	0	0	1	0	2	0	3
RPN2	1	2	2	1	2	0	8
SYA1	0	0	2	1	0	2	5
Total	9	4	10	6	5	2	37

* Values are given for each clade. The number of strains analyzed per clade is indicated.

Table 3

Departure from Hardy-Weinberg expectations for a collection of 150 *C. albicans* isolates

Clade	*ACC1*	*VPS13*	*GLN4*	*ADP1*	*RPN2*	*SYA1*	*6 loci*
I (52 isolates)	2+/0−	5+/0−	1+/0−	1+/0−	0+/0−	0+/0−	9+/0−
II (16 isolates)	0+/0−	1+/0−	0+/0−	0+/0−	5+/0−	5+/0−	11+/0−
III (24 isolates)	3+/0−	0+/0−	0+/0−	7+/0−	4+/0−	6+/0−	20+/0−
IV (11 isolates)	0+/0−	0+/0−	1+/0−	0+/0−	3+/0−	5+/0−	9+/0−
SA (21 isolates)	0+/0−	1+/0−	1+/0−	1+/0−	0+/0−	6+/0−	9+/0−
E (26 isolates)	0+/0−	3+/0−	0+/0−	8+/0−	3+/0−	2+/0−	16+/0−
Global (150 isolates)	0+/2−	1+/1−	0+/0−	0+/1−	0+/4−	0+/2−	1+/10−

The number of nucleotide positions with significant ($P<0.05$) excesses (+) or deficits (−) of heterozygotes are indicated for each locus in the six clades analyzed individually and when the same isolates were grouped together.

Table 4

Departure from Hardy-Weinberg expectations for a collection of 1044 isolates

Clade	*ACC1*	*ADP1*	*AAT1a*	*MPIb*	*SYA1*	*VPS13*	*ZWF1b*	*7 loci*	Polymorphic nucleotide sites[a]	
									Number	% excess heterozygotes
1 (467 isolates)	2+/0−	1+/0−	4+/1−	6+/0−	0+/0−	5+/0−	5+/0−	23+/1−	27	85.2%
2 (171 isolates)	1+/0−	1+/0−	0+/0−	0+/0−	5+/0−	1+/0−	6+/0−	14+/0−	20	70.0%
3 (117 isolates)	1+/0−	7+/0−	0+/1−	4+/0−	6+/0−	1+/0−	5+/0−	24+/1−	38	63.2%
4 (185 isolates)	0+/0−	1+/0−	0+/1−	0+/0−	6+/0−	1+/1−	2+/0−	10+/2−	20	50.0%
11 (104 isolates)	0+/0−	8+/0−	5+/0−	1+/0−	2+/0−	3+/0−	6+/0−	25+/0−	40	62.5%
Global analysis (1,044 isolates)	0+/4−	0+/1−	0+/3−	0+/5−	0+/6−	0+/4−	5+/2−	5+/25−	58	8.6%

The number of nucleotide positions with significant ($P<0.01$) excesses (+) or deficits (−) of heterozygotes are indicated at each locus for the five clades analyzed individually and when the same isolates were grouped together. Data used to assess departures from Hardy-Weinberg expectations for clades 1, 2, 3, 4, and 11 was from Odds *et al*. (2007). Cluster analysis of isolates studied with the MLST scheme used in the present study and the one used in Odds *et al.* (2207) demonstrated that clades 1, 2, 3, 4, and 11 defined by Odds *et al*. (2007) corresponded to clades I, II, III, SA, and E of the present study, respectively.

[a]The number of polymorphic nucleotide sites and the proportion of sites with significant excess heterozygotes is indicated for each clade. Nucleotide positions presenting a frequency > 0.95 (Supplemental Table 8) for the major allele were not considered to be significantly polymorphic and were not used in this analysis.

Fungal Genet Biol. Author manuscript; available in PMC 2009 March 1.