Uneven Distribution of Mating Types among Genotypes of *Candida glabrata* Isolates from Clinical Samples

Sylvain Brisse,1 Christine Pannier,2 Adela Angoulvant,3 Thierry de Meeus,4 Laure Dioncourt,1 Odile Faure,5 Héloïse Muller,6 Javier Peman,7 Maria Anna Viviani,8 Renée Grillot,5 Bernard Dujon,6 Cécile Fairhead,6 and Christophe Hennequin1,3,9

Institut Pasteur, Genotyping of Pathogens and Public Health, F-75015 Paris, France1; Laboratoire de Virologie, CHU Amiens, F-80000 Amiens, France2; Université Pierre et Marie Curie-Paris6, Laboratoire de Parasitologie-Mycologie, Paris, F-75012 France3; UMR 2724 IRD/CNRS “Génétique et Evolution des Maladies Infectieuses,” F-34000 Montpellier, France4; Laboratoire de Parasitologie-Mycologie, CHU Grenoble, F-38000 Grenoble, France5; Institut Pasteur, Génétique Moléculaire des Levures (URA 2171 CNRS, UFR 927 Université P. et M. Curie), F-75015 Paris, France6; Microbiology Department, La Fe University Hospital, Valencia, Spain7; Department of Public Health-Microbiology-Virology, Università degli Studi di Milano, Milan, Italy8; and INSERM, UMR511, Paris, F-75013 France9

Received 30 June 2008/Accepted 29 November 2008

In order to shed light on its basic biology, we initiated a population genetic analysis of *Candida glabrata*, an emerging pathogenic yeast with no sexual stage yet recognized. A worldwide collection of clinical strains was subjected to analysis using variable number of tandem repeats (VNTR) at nine loci. The clustering of strains obtained with this method was congruent with that obtained using sequence polymorphism of the *NMT1* gene, a locus previously proposed for lineage assignment. Linkage disequilibrium supported the hypothesis of a mainly clonal reproduction. No heterozygous diploid genotype was found. Minimum-spanning tree analysis of VNTR data revealed clonal expansions and associated genotypic diversification. Mating type analysis revealed that 80% of the strains examined are *MATa* and 20% *MATα* and that the two alleles are not evenly distributed. The *MATa* genotype dominated within large clonal groups that contained only one or a few *MATα* types. In contrast, two groups were dominated by *MATα* strains. Our data are consistent with rare independent mating type switching events occurring preferentially from type a to α, although the alternative possibility of selection favoring type a isolates cannot be excluded.

*Candida glabrata* has emerged as significant among the etiologic agents of invasive candidiasis and is of particular concern because of its natural resistance toazole derivatives (17, 39). This and/or the severity of underlying conditions of patients infected with *C. glabrata* may explain the higher mortality associated with this species compared to other *Candida* species (27). To date, knowledge about the basic biology and pathogenesis of *C. glabrata* remains limited in comparison with the amount of data accumulated for *Candida albicans*, the predominant pathogenic yeast. The population structure and mode of reproduction of a pathogen are highly relevant to questions such as those about the evolution of virulence and drug resistance or the relationship between sexual reproduction and virulence. Recent works on the population structure of *C. glabrata* concluded that reproduction is predominantly clonal, although the possibility of rare recombinant events has been suggested (9, 12). In contrast to *Saccharomyces cerevisiae*, *C. glabrata* is an obligate haploid which has apparently lost the ability to reproduce sexually, like many other fungal pathogens or symbionts (31).

As a part of a large comparative yeast genomics project, the complete genome of *C. glabrata* type strain ATCC 2001 (CBS138) was sequenced, assembled, and annotated (13). In contrast to the lack of an observed sexual stage, this and other work have shown that *C. glabrata* possesses three genetic cassettes, homologs of the *HML*, *HMR*, and *MAT* loci involved in mating type determination (*MATα* or *MATα*) in *S. cerevisiae* (36, 40). In *S. cerevisiae*, homothallism (self-fertility) is the result of a sophisticated mating type switching mechanism in which the *MAT* locus is converted from a to α and vice versa at each generation (25). This relies on a DNA double-strand break created by the HO endonuclease, followed by homologous recombination with one of the silent cassettes located in subtelomeric regions harboring either the a or α allele, namely, *HMR* and *HML*, respectively (37). Interestingly, in addition to an HO endonuclease gene homolog, *C. glabrata* also possesses homologs of the majority of genes involved in the mating and meiosis processes, as is the case for other fungi with no observed sexual cycle (31).

The presence of these genes in the genome does not, however, provide information on the existence and frequency of sexual reproduction in nature. Although the existence of both *MATα* and *MATα* types in natural populations of *C. glabrata* was described (5, 29), the distribution of mating types among genotypes and the possibility of mating type switching within populations has not been studied in depth.

Multilocus-based genotyping methods such as multilocus enzyme electrophoresis, multilocus sequence typing (MLST), and multilocus variable number of tandem repeats (VNTR) anal-
ysis are valuable tools for population structure analysis and phylogenetic reconstruction. Multi locus VNTR analysis allows improved discrimination (2), which can be useful for fine typing and short-term evolutionary questions, although phylogenetic relationships must be inferred with caution because of the homoplasy inherent to microsatellite markers (10). Length polymorphism analysis of such regions has been successfully applied for typing other fungi, such as C. albicans, Aspergillus fumigatus, Coccidioides immitis, and Saccharomyces cerevisiae (4). Like all eukaryotic genomes analyzed so far, the genome of C. glabrata reveals a substantial number of microsatellite sequences.

In this work, we have characterized by VNTR analysis the diversity and relationships at a microevolutionary level of a collection of 198 C. glabrata strains isolated mostly from clinical samples. Phylogenetic relationships at a deeper evolutionary level were assessed by NMT1 gene sequencing, while determination of the mating type was performed to investigate the evolutionary history of mat switching.

MATERIALS AND METHODS

Strains. One hundred ninety clinical isolates collected from distinct individuals and eight environmental or nonhuman isolates were tested (see Table S1 in the supplemental material). Strains had been isolated from Africa, Oceania, North and South America, and Europe from 1932 to 2004. One hundred thirty-seven strains had been isolated from Africa, Oceania, North and South America, and Europe from 1932 to 2004. One hundred thirty-seven strains had been isolated from Africa, Oceania, North and South America, and Europe from 1932 to 2004. One hundred thirty-seven strains had been isolated from Africa, Oceania, North and South America, and Europe from 1932 to 2004. One hundred thirty-seven strains had been isolated from Africa, Oceania, North and South America, and Europe from 1932 to 2004. One hundred thirty-seven strains had been isolated from Africa, Oceania, North and South America, and Europe from 1932 to 2004. One hundred thirty-seven strains had been isolated from Africa, Oceania, North and South America, and Europe from 1932 to 2004.

Microsatellite analysis. (i) Selection of loci. Version 20 of the assembly of the C. glabrata genome (type strain ATCC 2001) was used for the search of mononucle- mononucleotides using the Tandem Repeat Finder software from G. Benson (3) as implemented on Institut Pasteur’s web services (http://www.pasteur.fr/). Nine regions were selected, and primer pairs were designed and synthesized; the forward primers were labeled with a fluorophore (Applied Biosystems, Courtaboeuf, France) (Table 1).

(ii) Length polymorphism analysis. After a 2-day incubation of yeasts on Sabouraud agar plates (BioMérieux, Marcy l’Etoile, France), DNA was obtained from colonies using a one-step DNA extraction method with Chelex 100 (Sigma-Aldrich, St. Quentin-Fallavier, France) resin (24), adjusted to a 50-ng/μl concentration. Amplification was performed in a volume of 15 μl with 1× buffer, a 5 μM concentration of each oligonucleotide primer, 0.2 mM deoxynucleoside triphosphates (Pharmacia, Uppsala, Sweden), 0.6 U of Taq Gold DNA polymerase (Eurogentec, Liege, Belgium), and 1.2 μl of DNA. Thermal cycling parameters were as follows: 7 min at 94°C; 3 cycles of 30 s at 94°C, 30 s at 69°C, and 30 s at 72°C; 3 cycles of 30 s at 94°C, 30 s at 65°C, and 30 s at 72°C; 2 cycles of 30 s at 94°C, 30 s at 60°C, and 30 s at 72°C; 28 cycles of 30 s at 94°C, 30 s at 55°C, and 30 s at 72°C; and a final extension of 10 min at 72°C. PCR products for a given strain were then pooled together with a molecular marker (400HD-Rox; ABI) and sized in an automated sequencer (ABI377; ABI, Courtaboeuf, France). Sizes of the amplicons were automatically assigned using GenScan Analysis software (v3.1; ABI). Multilocus patterns were then determined using the Genotype vs.2 software (ABI).

Population genetic analysis. (i) Population structure based on VNTR and NMT1 sequence data. In a preliminary study of sequencing one example of several alleles observed for each locus (data not shown), we did not observe length changes associated with insertion/deletion within the flanking regions of the microsatellite loci. Consequently, fragment sizes were converted into deduced repeat sizes at each locus. Each unique allelic combination composed of the deduced number of repeats at the VNTR loci was considered a new type (RT). Phylogenetic relationships were estimated based on RT using the minimum-spanning tree (MStree) method as implemented in BioNumerics v5.10 (Applied-Maths, Sint-Martens-Latem, Belgium). The MStree algorithm links RTs so that the sum of the distances (number of distinct alleles between two RTs) is minimized (33); variants with the same number of mismatches to several RTs were preferentially linked to RTs with the highest total number of single-locus variants (SLVs), or double-locus variants in case of a match (16). Importantly, RT frequency was not taken as a criterion for determining links in cases of matches; hence, the founder genotypes of clonal groups were not identified by taking into account their frequency. In contrast to the unweighted-pair group method using average linkages (UPGMA) or phylogenetic methods, which place all extant genotypes (considered leaves) at the tips of the branches, in an MStree graph both the nodes and leaves correspond to observed genotypes (15). Therefore, in the graphical representation, strains with the same allelic profile fall in the same circle, whose size is proportional to the number of strains with that particular RT.

Homoplasy along the branches of the MStree was calculated using the Mul-tiLocus Analyzer program (S. Brisse, unpublished). The homoplasy was defined as 1 − (K − 1)/M, where K is the number of alleles and M the number of changes along the MStree. Hence, if each allele was generated by a single evolutionary change, the number of changes since the ancestral state would be K − 1, and the homoplasy index would be equal to zero. In contrast, if alleles are often changing by convergent evolution or reversion to an ancestral state, M would become much greater than K and the homoplasy index would increase toward 1. The same parameter was deduced along the UPGMA tree obtained based on profile data using PAUP* v4.38 (38).

### TABLE 1. Characteristics of the nine microsatellite-containing loci and respective oligonucleotide primers for amplification used for the analysis of length polymorphism

| Locus | Chromosome | Gene or intergene | Repeat in CBS138 | Forward primer | Reverse primer |
|-------|------------|-------------------|-----------------|----------------|----------------|
| 1     | F          | CAGLjF02849g YDR412w | (AAG)₇          | CGAGAGAAGAAATATGCT | CTTCAATTGGTTTACAGTC |
| 2     | I          | Intergene         | (GCT)₈         | GCACCTGTCTTTATATTAC | CGAATCCGTGATCCCTTC |
| 2bis  | D          | Intergene         | (AATC)₆        | ACCATAGGAAACCAACAA | TAGCCGTCATCCAGCATCA |
| 3     | A          | Intergene         | (ATC)₄         | ATTCGTTATTTAACAGTCC | ACCAAATGGAAACCAAG |
| 4     | H          | CAGLjH02783g, some similarities with Saccharomyces cerevisiae YJL076w NET1 | (ATC)₁₀ | TTATATCTTCTTGGGTGTCGG | CGAAAAACGTCAGAAACTC |
| 5     | J          | CAGLjJ010560g, similar to Saccharomyces cerevisiae YDR429w LRS4 | (ATC)₉          | CCTTTAAGGATGAGTACCTTC | GCTTGGTGTTAAAGGAAC |
| 6     | K          | CAGLjK05423g, similar to Saccharomyces cerevisiae YPR040w | (GTT)₁₃ | GTCTGGACCTTTGAATATGGTG | AATCGGTTAGTGTCCT |
| 8     | C          | CAGLjC01265g, weakly similar to Saccharomyces cerevisiae YIL115c NUP159 | (TAA)₁₁ | ATTTAAATATACAGCCTAC | CTGACGAACTTCTTCTT |
| 9     | I          | Intergene         | (ACTC)₃        | CCTCCCTCTGGCGGACT | TCACCTGGACCTCCCTGAG |

This table lists the characteristics of the nine microsatellite-containing loci and respective oligonucleotide primers for amplification used for the analysis of length polymorphism.
The sequence polymorphism of a fragment from the NMT1 gene (CAGLU0A04059a) has been shown to be sufficient for the assignment of C. glabrata strains into clades as previously defined by Dodgson et al. (11). The sequences of this 607-bp fragment in a selection of 148 strains were determined (see Table S1 in the supplemental material). New allele sequences were submitted to the MLST.net website (www.mlst.net). The distance matrices obtained with NMT1 typing and microsatellite typing were compared by use of a Mantel test using the Cadm software, which tests the null hypothesis of incongruence between matrices (28). One thousand permutations were performed, with an equal weight assigned to each matrix. A P value of <0.05 was considered significant.

The geographical structure of genetic variation was estimated by hierarchical F statistics at the levels of the continent (FCountry/Total), country (FCountry/Continental), and city (FCity/Continental) by Yang’s method (41) implemented in HierFstat v 0.04-4 (22) and tested with randomization of the appropriate subunits between units of the higher level to be tested. For instance, to test FCity/Continental, isolates were permuted across the different cities within countries, and FCountry/Continental was tested by permuting cities with all the isolates contained between the different countries within each continent. The statistic used was the G-based test as described by Goudet et al. (23) and was considered significant if the value was divided by 1,000. A simple description of HierFstat can be found in reference 7. Note that FCity/Continental could be assessed only for France and Brazil. Because a correlation between loci might bias population differentiation measures and testing, we repeated FST tests using the multilocus genotype of each individual as a unique locus with as many alleles as defined by the multilocus genotypes.

(ii) Mode of reproduction. In order to test the reproductive mode, the proportion of significant linkage disequilibrium among pairs of loci and the index of association were estimated. Linkage disequilibrium tests were implemented by Fstat 2.9.3.2. The statistic used is the log-likelihood ratio G. For each pair of loci, in each subsample, the G statistic is computed on each genotypic table of paired loci (28 pairs), and values are summed over all subsamples for each pair. For each pair of loci, in each subsample, genotypes are then randomly associated (free recombination), and the G statistic is computed and then summed over all subsamples for this randomized data set. This randomization was repeated 10,000 times. The P value obtained was the number of times a randomized G was as high as or higher than the one measured in the real data set (7, 8). Because repetitive tests enhance type I error, we also corrected these P values using the sequential Bonferroni procedure (26), where P values are ranked from the lowest to the highest and the first one is multiplied by the total number of tests (n = 28), the second by n − 1, and so on until the corrected P values reaches 1. Multilocus linkage was measured with Agapow and Burt’s r2 (1) in each subsample and its significance tested by 1,000 randomizations (random reassociation of alleles at the different loci) with Multilocus v 1.3b (available at http://www.agapow.net/software/multilocus/ and updated from reference 1). This provided a series of independent tests that were combined by the Fisher (20) procedure in which the n P values of a test series are combined as χ² = −2 ∑ ln(Pi), a quantity that can be compared to a chi-square distribution with 2n degrees of freedom.

In case of strong linkage disequilibrium, using each locus as an independent unit of information for population structure analyses may lead to erroneous conclusions. We thus reanalyzed all our results by using the multilocus genotype as a single locus and each RT as a different allele at that locus.

Comparison of means was undertaken with a Wilcoxon rank sum test and comparison of variances with a F test (34) with S-Plus 2000 Professional release 3 (MathSoft Inc., 2000).

Mating type determination. Based on the sequence of the type strain ATCC 2001 (CBS138) (13) and analyzed their length polymorphism in 198 C. glabrata isolates. For the nine tested loci, all strains tested gave a unique PCR product, the sizes of which were used to deduce the repeat sizes, as indicated in Table S1 in the supplemental material and available through the MLVA-NET website data system at www.pasteur.fr/mlva. Locus 1 did not exhibit any polymorphism among the 198 tested strains and was excluded from further analyses. For the other loci, three to eight different alleles were found per locus, with a total of 41 different alleles. The combination of alleles from the eight loci generated 90 different genotypes, or RTs, numbered RT1 to RT90. The four most frequent RTs were RT15 (n = 28 strains), RT13 (n = 19), RT52 (n = 13), and RT41 (n = 11). A total of 63 RTs were represented by a single strain. Overall, the genetic diversity (probability that two strains taken at random have different genotypes) was 0.96. When the eight loci were incorporated successively in the analysis, the genetic diversity reached a plateau, indicating that testing more loci would not significantly enhance the discriminatory power (not shown).

Population structure of C. glabrata deduced from VNTR data. In order to visualize the genetic diversity, an MSTree analysis of the 90 distinct RTs was performed (Fig. 1). In this graphical representation, strains with the same allelic profile fall in the same circle, the size of which is proportional to the number of strains with that particular RT. We observe that most links correspond to only one allelic mismatch and could therefore be considered as corresponding to single evolutionary steps. Such pairs of RTs are defined as SLVs. However, the homoplasy index calculated along the MSTree was 0.73 (range, 0.67 to 0.81 per locus), indicating that inferred phylogenetic links between genotypes that are not very closely related are probably unreliable. Similarly, the homoplasy index calculated along the UPGMA dendrogram (not shown) was 0.78.

The mode of reproduction was tested by the calculation of linkage disequilibrium between loci, which is a major criterion for testing clonality of haploid microorganisms. Within countries, among the 28 possible pairs of loci, 23 (82%) were in significant linkage disequilibrium, with each locus being in significant linkage with several other loci. At the sequential Bonferroni level, 19 pairs (68%) remained significant and all loci were involved several times in a significant linkage, except for vntr9, which was never found statistically associated with any locus at this level; the latter result may be explained by the slightly increased homoplasy level of vntr9 (0.81) compared to the other loci (range, 0.67 to 0.79). Differentiation was met between countries but not at smaller scales and not between different dates. Consequently, for linkage disequilibrium analyses, all strains from each country were considered as subsamples. Multilocus tests provided six subsamples, out of eight where the test was possible, with a significant linkage (τp > 0. P < 0.005). Overall, the mean multilocus linkage (weighted for sample size) was τp = 0.155, with an associated combined P value of <0.001 (obtained with Fisher’s procedure). Overall, these results indicate a strong and global (genomic) association between all loci in the different subsamples. Closer examination of the MSTree shows that the most frequent RTs (the largest circles in Fig. 1) were often placed in a central position, indicating that testing more loci would not significantly enhance the discriminatory power (not shown).

RESULTS

VNTR-based genotypic diversity of C. glabrata. We selected nine microsatellite regions from the complete sequence of the type strain ATCC 2001 (CBS138) (13) and analyzed their

---

**Table S1** in the supplemental material and available through the MLVA-NET website data system at www.pasteur.fr/mlva.

---

**Figure 1** in the article. A graphical representation of the genetic diversity among 198 C. glabrata isolates. The x-axis represents different alleles, and the y-axis represents different RTs. Each RT is represented by a circle, with the size of the circle proportional to the number of strains that have that RT. The figure illustrates the strong population structure and the presence of multiple alleles and RTs. The red lines indicate significant linkage disequilibria, while the blue lines indicate non-significant linkages.
eration of a number of variants, each being more closely related to the founder genotype than to other variants. Thus, each central frequent RT and its associated SLVs may be considered as descending from a single ancestor. Given the high homoplasy index, the progressive diversification of two independent expanding clones may result in similar or identical genotypes by evolutionary convergence. Therefore, we chose to restrict the definition of clonal complexes (CCs) to the founder RTs with at least four isolates and their SLVs. With this definition, there were six CCs, which were named according to the number of the founder RT (CC6, CC15, CC41, CC52, CC64, and CC77), as shown in Fig. 1. Altogether, these CCs include 143 strains out of 198 (72%). It must be noted that in CC15 and CC52, one SLV itself included more than four strains (RT13 and RT48, respectively), and therefore its own SLVs were also included in the complex. CC77 was a particular case, as it had no SLV.

**Correspondence of VNTR data with NMT1 sequences.** In order to establish the correspondence of the six CCs with phylogenetic groups previously defined by MLST (12) and to define the phylogenetic relationships among the six CCs, an internal portion of the NMT1 gene in representative strains of each CC and other strains was sequenced (see Table S1 in the supplemental material). Indeed, alleles of this gene were previously shown to be strongly associated with MLST groups (11). Twenty distinct NMT1 alleles were found. In addition to eight alleles already described, new allele sequences were found, which were referenced as NMT1-35 to NMT1-39 (http://cglabrata.mlst.net). NMT1 alleles were mapped on the MStree (Fig. 1). Remarkably, the NMT1 allele distribution was highly concordant with the VNTR data. Indeed, CC6, CC15, CC41, CC52, CC64, and CC77 were each characterized by a predominant NMT1 allele (NMT1-2, NMT1-8, NMT1-3, NMT1-6, NMT1-16, and NMT1-4, respectively). These alleles had previously been shown to characterize groups II, I, III, ST22, ST25, and IV, respectively (12). One or a few isolates of four of these CCs (CC6, CC15, CC52, and CC64) contained variant NMT1 alleles. Examination of these allele pairs (NMT1-5 and NMT1-32, NMT1-8 and NMT1-39, NMT1-6 and NMT1-37, NMT1-6 and NMT1-35, and NMT1-2 and NMT1-3) showed that variant alleles differed by only one single-nucleotide polymorphism from the dominant allele of their respective CC, with the exception of NMT1-2 and NMT1-3 in CC6, with three single-nucleotide polymorphisms between them (see Fig. 1). Alleles from the strains sequenced outside of CCs
were generally distinct from those characterizing the CCs, with a few exceptions possibly attributable to homoplasy or to our restricted definition of CCs (Fig. 1). The correspondence between VNTR and \( \text{NMT1} \) was further supported by the high level of significance of the Mantel test comparing the two distance matrices (\( P < 0.01 \)).

**Geographical and temporal distribution of isolates.** The influence of geographical origin on the genotype distribution was tested at three hierarchical levels, i.e., continent, country, and city. Hierarchical analyses were undertaken for all isolates sampled between 1993 and 2003 with at least two representatives per city. Subsamples with no date, older than 1993, or with a single isolate were removed. We observed a weak and nonsignificant differentiation between cities within countries (\( F_{\text{City/Country}} = 0.012, P = 0.315 \)), which is confirmed by the multilocus genotype analysis (\( F_{\text{City/Country}} = 0.007, P = 0.486 \)).

Between countries, differentiation is medium but highly significant with the eight microsatellite loci or multilocus genotypes (\( F_{\text{Country/Continent}} = 0.07, P = 0.001 \) in both cases). With microsatellites, differentiation between continents is the strongest (\( F_{\text{Continent/Total}} = 0.11, P = 0.015 \)), while multilocus genotypes provide here a contradictory result (\( F_{\text{Continent/Total}} = -0.03, P = 0.415 \)). This geographic structure is reflected by the uneven distribution of some RTs or CCs. For example, although the majority of strains were isolated in Europe, CC6 and CC64 show a high proportion of isolates from North America (see Table S1 in the supplemental material). Only RT15 was isolated from four different continents. In contrast, the group of related RTs, RT72 to RT75, originated solely from North America.

Our strain collection was isolated from patients between the years 1996 and 2003, in addition to six reference strains isolated between 1932 and 1988. A high number of RTs and all six defined CCs were isolated across several years. One extreme case was RT15, isolated over a span of 40 years (see Table S1 in the supplemental material), which suggests that VNTR profiles can be stable over this time scale.

**Phylogenetic distribution of mating type.** We determined the mating types of 162 strains by a four-PCR system specifically targeting the genes at the \( \text{MAT} \) locus (Fig. 2). Nearly 80\% (\( n = 132 \)) of our strains corresponded to \( \text{MATa} \) and 20\% (\( n = 30 \)) to \( \text{MATa} \) (see Table S1 in the supplemental material).

There was no difference in the distribution of mating type according to either continent or country. Nevertheless, the frequency of multilocus genotypes defined by VNTR data appeared to be unevenly distributed between the two mating types. Mating type \( a \) was more often associated with repeated genotypes (mean = 2.42) than mating type \( \alpha \) (mean = 1.50), although the difference was not statistically significant (Wilcoxon rank sum test, \( P = 0.602 \)). However, the variance in genotype frequency of each mating type was very different (14.17 for \( a \) and 1 for \( \alpha \); \( F \) test, \( P < 0.001 \)). This is due to the fact that the \( a \) type was often associated with highly frequent RTs, while \( \alpha \) was never associated with RTs encountered more than four times.

In order to establish the phylogenetic relationships of a number of strains of both mating types, we used the \( \text{NMT1} \) sequences to deduce the long-term phylogenetic relationships and used the VNTR data to infer microevolution. Since there was a one-to-one correspondence between CCs and the branches in the \( \text{NMT1} \) phylogeny, we placed the VNTR-based CCs in front of their corresponding branch (Fig. 3). Clearly, many well-supported \( \text{NMT1} \) branches contained both mating types. In four of the CCs, strains were predominantly \( \text{MATa} \). Nevertheless, each of these contained at least one isolate of mating type \( \alpha \). The phylogenetic association of these rare \( \text{MATa} \) isolates with \( \text{MATa} \) isolates, initially deduced based on multilocus VNTR data, was thus fully confirmed by \( \text{NMT1} \) gene sequencing. Contrasting with the predominance of \( \text{MATa} \) in other CCs, CC64 and CC77 comprised 18 \( \text{MATa} \) strains out of 30 total. Remarkably, CC64 was made up exclusively of \( \text{MATa} \) strains. CC64 and CC77 did not appear to be associated together in a single branch based on \( \text{NMT1} \) phylogeny, suggesting an independent origin of the \( \alpha \) type in these two groups.

**DISCUSSION**

Population genetic studies of \( C. \) *glabrata* are of major importance to understand the biology and evolution of this prominent emerging pathogen among the etiologic agents of candidiasis. We defined VNTR markers based on the entire genome sequence of strain ATCC 2001 and have used these markers to analyze a collection of 198 \( C. \) *glabrata* strains. VNTR typing is a powerful tool for fine-scale population genetic studies and has been used for a number of fungal pathogens, such as \( \text{Coccidioides immitis} \), \( \text{Penicillium marneffei} \), and \( \text{C. albicans} \) (18, 19, 21).

The VNTR data were used to build an MStree that enabled us to define subpopulations, or CCs. The six CCs were not clearly demarcated based on VNTR data, as several of them were connected on the MStree by a succession of links corresponding to single allelic mismatches. This was in sharp contrast to the \( \text{NMT1} \)-based phylogeny, in which all CCs appeared to be very distinct, with restricted amounts of sequence polymorphism within CCs, compared to sequence divergence among CCs (Fig. 3). We attribute these contrasting results to homoplasy in the VNTR data, which was very high. In contrast, homoplasy was virtually absent within CCs. Therefore, VNTR and nucleotide sequences in housekeeping genes clearly appear as complementary genetic markers that are suitable at distinct evolutionary time scales. Nevertheless, concordance between VNTR-based groupings and \( \text{NMT1} \) sequences supports the idea that the disclosed CCs represent distinctive genetic entities. \( \text{NMT1} \) sequences were previously shown to correspond strongly to phylogenetic groups, but two new groups were defined by our analysis, as was also the case in a recent study (29), highlighting the importance of representative strain sampling in order to achieve an accurate description of \( C. \) *glabrata* strain diversity.

It was interesting to observe that the central genotypes of CCs, as identified by MStree analysis, were also the most frequent ones. It is important to note that the central genotypes were identified by the MStree algorithm without taking into account their frequency in the population but based only on the number of their SLVs. Therefore, founder genotypes (RT52, RT41, RT15, and RT13) appear to represent the natural population of “infective” strains better than laboratory reference strains, which may have diverged genetically, perhaps because of nonoptimal modes of preservation in labora-
tories. For example, one can notice that the type strain (ATCC 2001), which was used for the complete genome sequence project, is not included in any major CC and thus may not be the best representative strain of the species. Further attention should be given to these aspects with respect to virulence experiments. In addition, since biological differences such as resistance to flucytosine or alleles of the ALS genes (32, 35) have been previously demonstrated for genetic subgroups of C. albicans, differences in clinically relevant properties should also be evaluated for C. glabrata CCs or RTs. Selecting representatives of each of the main CCs for the generalization of any observation to the entire species is thus recommended.

The possible existence of mating and its impact at the population level were also investigated. Many pathogenic hemiascomycetes, such as C. albicans, Candida tropicalis, and Candida parapsilosis, are considered to reproduce asexually most of the time (31). In contrast to the case for MLST analysis, incompatibility between loci cannot be tested with VNTR data because of homoplasy inherent to the method (10). However, the strong linkage disequilibrium between alleles for all pairs of loci and the significant index of association demonstrated in this study support the hypothesis of a mainly clonal mode of reproduction in C. glabrata. Strong association of NMT1 alleles with VNTR-based CCs also suggests that homologous recombination of NMT1 alleles does not cause frequent replacement of alleles among distinct genetic backgrounds. Of note, we never observed the simultaneous presence of two VNTR alleles. This could be explained either by an absence of mating, by an instability of diploids, or by the presence of homozygote diploids resulting from inbreeding. Nonetheless, these homozygote diploids could still be expected to be heterozygous at the MAT locus, a situation we never encountered in our mating type analysis. The pattern observed for several CCs, with a predominant genotype giving rise to closely related genotypes, is concordant with the clonal expansion-diversification model of microbial evolution (15), according to which successful ge-

FIG. 2. Mating type determination for C. glabrata strains. (A) Genomic organization of Mat cassettes in C. glabrata and primers designed to specifically amplify genes at the MAT locus. Terminology for the HML, HMR, and MAT loci was by homology with S. cerevisiae. Note that the HML and MAT loci are located on chromosome B, while the HMR locus is located on chromosome E. (B) Results of amplification targeting loci α1 (upper left), α2 (upper right), a1 (lower left), and a2 (lower right). Type α strains (lanes 5, 7, and 8) are US02Bal12, EF1521Blo1, and EF0510Blo1. Type a strains (lanes 1 to 4, 6, 9, and 10) are EF1515Blo1 ES0085Blo1, US003NY110, US02Bal020, EF1213Blo1, EF1204Blo1, EI4101Blo1, EF1031Blo1, US02Bal003, and USTCC90030.
notypes generate clonal families by multiple independent diversifications from an ancestral genotype. Overall, these results are in accordance with two previous studies, using multilocus enzyme electrophoresis and MLST, that concluded that the mode of reproduction of C. glabrata was mainly clonal (9, 12). Note that a predominantly clonal mode of reproduction is not incompatible with rare events of genetic recombination, which were reported (12). The fact that the central genotype is also found to be the most frequent in the CC implies that the clonal expansion is recent or still going on. Hence, the data suggest that as some genotypes (RTs) increase in frequency, they diversify by giving rise to a number of variants, which can be revealed by VNTR analysis.

Even though no evidence for mating has ever been found in this species, opposite mating types have been detected. In our collection of strains, we found 80% MATa cells. This predominance of MATa strains is in accordance with the results reported by Srikantha et al. (36), who found 58% MATa strains among 39 tested strains. In contrast, Lin et al. found only 32% MATa strains in a panel collected from 37 patients, but since all of these originated from Taiwan, this result may reflect differences in prevalence of mating types and/or clones according to geography (29).

In the present study, the distribution of mating types among genotypes revealed two kinds of asymmetry. First, in most CCs there were few MATa/H9251 cells, whereas more than half of the total MATa cells were clustered in two particular CCs. Hence, the distribution of mating types is highly nonrandom. This observation seems incompatible with switching at frequencies in the range of VNTR marker evolutionary rate, unless selection is acting against MATa strains. Second, MATa was associated with rare genotypes (involving fewer than five isolates), in contrast to MATa, which dominated frequent genotypes (up to 22 isolates per RT). This result may be explained by a selective advantage of MATa isolates in a clinical context or simply by the increased exposure of patients to MATa isolates. Regard-
ing the phylogenetic distribution of mating types, our data support the hypothesis that mating type switching occurred several times independently, at least once in each of CC6, CC15, CC41, CC52, and CC77. We observed the existence of few MATα strains in the MATα-dominated CC6, CC15, CC41, and CC52. At first sight, this observation could suggest a preferential directionality of mating type switching from MATα to MATα, rather than the reverse. Under this hypothesis, the maintenance of MATα isolates in the population could be due to a selective advantage. However, it is not possible to exclude the possibility that switching occurred from α to a in these groups, followed by a selective advantage of MATα isolates. Nevertheless, the fact that the MATα-dominated CC64 did not show a single MATα variant, even if it had time to diversify into several distinct RTs, could reflect the fact that switching from α to a is rare, at least in the CC64 genomic background. Indeed, we believe previous reports support our data: in the laboratory, switching has been reported in PCR experiments but never in live cells despite intensive searching (5, 6; our unpublished results). In human patients, switching is mentioned in two studies, where isogenic strains of both mating types were isolated from a single patient. In one study, it was speculated that switching from a to α occurred (5), and in the second, an MATα isolate was collected later than two MATα isolates (29; C. Y. Lin, personal communication).

In S. cerevisiae, where switching occurs every other generation, it is driven by a DNA cut by the Ho endonuclease at the MAT locus. While the C. glabrata HO homolog does not contain any of the point mutations known to inactivate HO in S. cerevisiae (6, 14), and while this gene is transcribed in growing cells (30), it is not known whether the enzyme is active, in which case its regulation could still be very different from that in S. cerevisiae. Even in the absence of an active HO, switching could occur by gene conversion between the MAT locus and the HML and HMR loci. Most C. glabrata cells harbor type α information at HML and type a information at HMR, like S. cerevisiae. The genome sequence has revealed that HML and MAT are located on the same chromosome, while HMR is located on a different chromosome (14). This may explain the more frequent occurrence of switching from MATα to MATα, while the reverse switch would involve different chromosomal and may be more difficult to achieve, thus providing a molecular hypothesis for the asymmetry in mating type distribution reported here. Alternately, our sample, which is dominated by clinical strains, may be biased in favor of strains of type MATα, while type MATα strains may predominate in other environments (that remains to be described). It is conceivable that α strains have a selective disadvantage in humans or are less pathogenic and hence less frequently retrieved from clinical samples. It is noteworthy that switching is not necessarily indicative of sexual reproduction. For mating to occur, cells of opposite mating types must meet, and this may be rare given the single-clone dominance found in infected patients. Mating type switching in a single cell would theoretically provide a possibility for mating between sister cells. This warrants further studies of the possible link between mating and successful evolution of C. glabrata strains as human pathogens, including systematic MAT typing during the course of infection in patients at different body sites.

ACKNOWLEDGMENTS

This work was supported in part by ACI no. MIC0314 and GDR 2354 from the CNRS. The platform Genotyping of Pathogens and Public Health receives financial support from Institut Pasteur and the Institut de Veille Sanitaire. T. De Meeus is supported by the CNRS and IRD.

We thank colleagues from the Unité de Génétique Moléculaire des Levures and from Génolevures for stimulating discussions. We are indebted to Bill Merz (Baltimore, MD), D. Perlin (Newark, NJ), and G. Sanson (Sao Paulo, Brazil), who provided some strains tested in this work.

REFERENCES

1. Agapow, P. M., and A. Burt. 2001. Indices of multilocus linkage disequilibrium. Mol. Ecol. Notes 1:101–102.
2. Botterel, F., C. Desterke, C. Costa, and S. Bretagne. 2001. Analysis of microsatellite markers of Candida albicans used for rapid typing. J. Clin. Microbiol. 39:4076–4081.
3. Broekz, P. J., S. A. Lachke, T. Srikanta, C. Pujol, R. Galask, and D. R. Soll. 2003. Phenotypic switching and mating type switching of Candida glabrata at sites of colonization. Infect. Immun. 71:1709–1718.
4. Butler, G., C. Kenny, A. Fagan, C. Kurischko, C. Gaillardin, and K. H. Waller. 2004. Evolution of the MAT loci of Candida glabrata in yeast species. Proc. Natl. Acad. Sci. USA 101:1632–1637.
5. De Meeus, T., and J. Goudet. 2007. A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. Infect. Genet. Evol. 7:731–735.
6. De Meeus, T., K. D. McCoy, P. Prugnolle, C. Chevillon, P. Durand, S. Hurtrez-Boussès, and F. Renaud. 2007. Population genetics and molecular epidemiology or how to “debusquer la bête.” Infect. Genet. Evol. 7:306–332.
7. De Meeus, T., F. Renaud, E. Mouveroux, J. Reyes, G. Gakeazzi, M. Maille, and J. M. Ristide. 2002. Genetic structure of Candida albicans populations in AIDS and non-AIDS patients. J. Clin. Microbiol. 40:2199–2206.
8. Dettman, J. R., and J. W. Taylor. 2004. Mutation and evolution of microsatellite loci in Neurospora. Genetics 166:1231–1240.
9. Dodson, A. R., C. Pujol, D. W. Denning, D. R. Soll, and A. J. Fox. 2003. Multilocus sequence typing of Candida glabrata reveals geographically enriched clades. J. Clin. Microbiol. 41:5709–5717.
10. Dodson, A. R., C. Pujol, M. A. Pfaffer, D. W. Denning, and D. R. Soll. 2005. Evidence for recombination in Candida glabrata. Fungal Genet. Biol. 42:233–243.
11. Dujon, B., D. Sherman, G. Fischer, P. Durrens, S. Casaregola, I. Lafontaine, J. De Montigny, C. March, C. Neugee, E. Talla, N. Goffard, L. Frangeul, M. Aygi, V. Ailhaud, A. Baud, V. Barbe, S. Barnay, S. Blanchin, J. M. Beckerich, E. Beine, C. Bleykasten, A. Boisram, J. Boyer, L. Cattolico, F. Confanioliero, A. D. Darbau, L. Despos, E. Fabre, C. Fairhead, H. Ferry-Dumazet, A. Gralli, F. van Huyte, C. Hennecou, N. Maimains, P. Joyet, R. Kager, B. M. Lacke, M. Lemaire, I. Lesur, M. Ma, H. Muller, J. M. Nicaud, M. Nikolski, S. Oztas, O. Ozier-Kalogeropoulos, S. Pellenz, S. Poiteur, G. F. Richard, M. L. Straub, A. Suleau, D. Swennen, F. Tekia, M. Wesolowski-Louvall, E. Westholf, B. Wirth, M. Zenou-Meyer, I. Zivanovic, M. Bolotin-Pukhalara, A. Thierry, C. Bouchier, B. Cournon, C. Scarrelli, C. Gaillardin, J. Weissbach, P. Winkler, and J. L. Souciet. 2004. Genome evolution in yeasts. Nature 430:35–44.
12. Fabre, E., H. Muller, P. Thirioz, I. Lafontaine, B. Dujon, and C. Fairhead. 2005. Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and subtelomeres. Mol. Biol. Evol. 22:585–585.
13. Feil, E. J. 2004. Small change: keeping pace with microevolution. Nat. Rev. Microbiol. 2:495–505.
14. Fabre, E., J. L. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. 186:1518–1530.
15. Fidler, P. L., J. A. Vazquez, and J. D. Sobel. 1999. Candida glabrata: review of epidemiology, pathogenesis, and clinical disease with comparison to C. albicans. Clin. Microbiol. Rev. 12:80–96.
16. Fisher, M. C., D. Aanensen, S. de Hoog, and N. Vannitakom. 2004. Multilocus microsatellite typing system for Penicillium marneffei reveals spatially structured populations. J. Clin. Microbiol. 42:5056–5069.
17. Fisher, M. C., G. L. Koenig, T. J. White, G. San-Bas, R. Negroni, I. G. Alvarez, B. Vanke, and J. W. Taylor. 2001. Biogeographic range expansion into South America by Coccioides immitis mirrors New World patterns of human migration. Proc. Natl. Acad. Sci. USA 98:4558–4562.
18. Fisher, R. A. 1970. Statistical methods for research workers, 14th ed. Oliver and Boyd, Edinburgh, United Kingdom.
21. Fundyg, R. E., T. J. Lott, and J. Arnold. 2002. Population structure of Candida albicans, a member of the human flora, as determined by microsatellite loci. Infect. Genet. Evol. 2:57–68.

22. Goudet, J. 2005. Hierfstat, a package for R to compute and test hierarchical F-statistics. Mol. Ecol. Notes 5:184–186.

23. Goudet, J., M. Raymond, T. De Meûs, and F. Rousset. 1996. Testing differentiation in diploid populations. Genetics 144:1933–1940.

24. Hennequin, C., E. Abachin, F. Symoens, V. Lavarde, G. Reboux, N. Nolard, and P. Berche. 1999. Identification of Fusarium species involved in human infections by 26S rRNA gene sequencing. J. Clin. Microbiol. 37:3586–3589.

25. Herskowitz, I. 1988. Life cycle of the budding yeast Saccharomyces cerevisiae. Microbiol. Rev. 52:536–553.

26. Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6:65–70.

27. Klevay, M. J., E. J. Ernst, J. L. Hollanbaugh, J. G. Miller, M. A. Pfaller, and D. J. Diekema. 2008. Therapy and outcome of Candida glabrata versus Candida albicans bloodstream infection. Diagn. Microbiol. Infect. Dis. 60:273–277.

28. Legendre, P., and F. Lapointe. 2004. Assessing the congruence among distance matrices; single malt Scotch whiskies revisited. Aust. New Zealand J. Stat. 46:615–629.

29. Lin, C. Y., Y. C. Chen, H. J. Lo, K. W. Chen, and S. Y. Li. 2007. Assessment of Candida glabrata strain relatedness by pulsed-field gel electrophoresis and multilocus sequence typing. J. Clin. Microbiol. 45:2452–2459.

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

31. Nielsen, K., and J. Heitman. 2007. Sex and virulence of human pathogenic fungi. Adv. Genet. 57:143–173.

32. Pujol, C., M. A. Pfaller, and D. R. Soll. 2004. Fluconazole resistance is restricted to a single genetic clade of Candida albicans. Antimicrob. Agents Chemother. 48:262–266.

33. Schouls, L. M., H. G. van der Heide, L. Vauterin, P. Vauterin, and F. R. Mool. 2004. Multiple-locus variable-number tandem repeat analysis of Dutch Botrytella paxillensis strains reveals rapid genetic changes with clonal expansion during the late 1990s. J. Bacteriol. 186:5496–5505.

34. Sokal, R. R., and F. J. Rohlf. 1981. Biometry, 2nd ed. Freeman and Co., New York, NY.

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

36. Srikantha, T., S. A. Lachke, and D. R. Soll. 2003. Three mating type-like loci in Candida glabrata. Eukaryot. Cell 2:328–340.

37. Strathern, J. N., A. J. Klar, J. B. Hicks, J. A. Abraham, J. M. Ivy, K. A. Nasmyth, and C. McGill. 1982. Homothallic switching of yeast mating type cassettes is initiated by a double-stranded cut in the MAT locus. Cell 31:183–192.

38. Swofford, D. L. 1988. PAUP*, phylogenetic analysis using parsimony (* and other methods), 4th ed. Sinauer, Sunderland, MA.

39. Trick, W. E., S. K. Fridkin, J. R. Edwards, R. A. Hajjeh, and R. P. Gaynes. 2002. Secular trend of hospital-acquired candidemia among intensive care unit patients in the United States during 1989–1999. Clin. Infect. Dis. 35:627–630.

40. Wong, S., M. A. Fares, W. Zimmermann, G. Butler, and K. H. Wolfe. 2003. Evidence from comparative genomics for a complete sexual cycle in the ‘asexual’ pathogenic yeast Candida glabrata. Genome Biol. 4:R10.

41. Yang, R. C. 1998. Estimating hierarchical $F$-statistics. Evolution 52:950–956.