

## Optimization and Validation of Multilocus Sequence Typing for *Candida albicans*

Arianna Tavanti,<sup>1,2</sup> Neil A. R. Gow,<sup>1</sup> Sonia Senesi,<sup>2</sup> Martin C. J. Maiden,<sup>3</sup>  
and Frank C. Odds<sup>1\*</sup>

Department of Molecular & Cell Biology, Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD,<sup>1</sup> and Peter Medawar Building for Pathogen Research and Department of Zoology, University of Oxford, Oxford OX1 3SY,<sup>3</sup> United Kingdom, and Dipartimento di Patologia Sperimentale, Biotechnologie Mediche, Infettivologia ed Epidemiologia, Università degli Studi di Pisa, Pisa 56127, Italy<sup>2</sup>

Received 20 March 2003/Returned for modification 5 May 2003/Accepted 7 May 2003

**Multilocus sequence typing (MLST) was applied to 75 *Candida albicans* isolates, including 2 that were expected to be identical, 48 that came from diverse geographical and clinical sources, and 15 that were sequential isolates from two patients. DNA fragments ( $\approx 500$  bp) of eight genes encoding housekeeping functions were sequenced, including four that have been described before for *C. albicans* MLST, and four new gene fragments, *AAT1a*, *AAT1b*, *MPI*, and *ZWF1*. In total, 87 polymorphic sites were found among 50 notionally different isolates, giving 46 unique sequence types, underlining the power of MLST to differentiate isolates for epidemiological studies. Additional typing information was obtained by detecting variations in size at the transcribed spacer region of the 25S rRNA gene and tests for homozygosity at the mating type-like (MTL) locus. The stability of MLST was confirmed in two sets of consecutive isolates from two patients. In each set the isolates were identical or varied by a single nucleotide. Reference strain SC5314 and a derived mutant, CAF2, gave identical MLST types. Heterozygous polymorphisms were found in at least one isolate for all but 16 (18.4%) of the variable nucleotides, and 35 (41%) of the 87 individual sequence changes generated nonsynonymous amino acids. Cloning and restriction digestion of a gene fragment containing heterozygous polymorphisms indicated that the heterozygosity was genuine and not the result of sequencing errors. Our data validate and extend previous MLST results for *C. albicans*, and we propose an optimized system based on sequencing eight gene fragments for routine MLST with this species.**

The advent of high-throughput nucleotide sequencing technology has provided numerous opportunities for routine and unambiguous microbial isolate characterization. In addition to providing accurate and portable information, techniques such as multilocus sequence typing (MLST) (19) permit epidemiological isolate characterization to be integrated with population and evolutionary studies. MLST is a generic technique that has been exploited for several bacterial species (5, 7, 8, 23), and a number of web-accessible databases are available that enable the rapid dissemination and comparison of isolate characterization data (<http://www.mlst.net>).

The particular advantages of MLST as a typing method are that DNA nucleotide sequences can be determined by automated technology with minimal subjective interpretation of data such as exists in all methods dependent on phenotypic characteristics, fermentation profiles, and other qualitative comparators. In addition, MLST data from different sources can be archived and distributed electronically and interrogated and added to from distant locations to facilitate comparisons for global epidemiology and population studies. A web site (<http://www.mlst.net/new/index.htm>) has already been created for data archiving and analysis with six pathogenic bacteria.

Recently, Bougnoux et al. described an MLST system for the opportunistic fungal pathogen *Candida albicans* (2). Many ap-

proaches to strain typing have been developed for this species (32), but no system has yet achieved universal acceptance. Those based on *C. albicans* nucleotide sequences show greater or lesser diversity of types depending on the extent of conservation of the target sequences. Typing based on the intergenic transcribed spacer sequences of genes encoding rRNA tend to differentiate isolates into a very small number of major subclasses (22, 33), and similar conserved sequences have also been used for differentiation at the species level (4). By contrast, DNA fingerprints revealed with oligonucleotide probes for widely dispersed repeat sequences in the *C. albicans* genome show a great diversity of strain types (28, 29) and are even capable of revealing minor genomic adaptations to host microenvironments, a process known as microevolution (25, 31). MLST, based on allelic variation in the nonconserved portion of unrelated genes, aims to provide characterization that is sufficiently conservative to be robust and reproducible but provides levels of discrimination appropriate for the purposes both of investigation of clinically relevant problems, such as epidemic outbreaks of infection and resistance to antifungal agents, and of population analyses.

Most MLST schemes described to date have been for haploid microorganisms. The permanently diploid chromosome complement in *C. albicans* allows extra differentiation of isolates in this species because MLST data for some isolates may show two bases at the same variable site, indicative of the presence of two diploid alleles in these diploid organisms (2).

To optimize and validate MLST for typing *C. albicans*, we analyzed results with 75 *C. albicans* isolates for sequences of

\* Corresponding author. Mailing address: Department of Molecular & Cell Biology, Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, United Kingdom. Phone and fax: (44) 1224 273128. E-mail: f.odds@abdn.ac.uk.

TABLE 1. List of gene fragments used for *C. albicans* MLST, with details of primers

Fragment	Gene product	Primer	Amplicon size (bp)	5' sequence start	3' sequence start	Sequenced fragment (bp)	No. of variable sites	
							Nucleotides	Amino acids
<i>AAT1a</i>	Aspartate aminotransferase	Fwd 5'-ACTCAAGCTAGATTTTGGC-3' Rev 5'-CAGCAACATGATTAAGCCC-3'	478	ATTGAA	CGTTTC	373	7	1
<i>AAT1b</i>	Aspartate aminotransferase	Fwd 5'-ATGGCTTATCAAGTTTTC-3' Rev 5'-GTGGCATAAATGAATATTCG-3'	491	TTGACTAA	CTGGGACCA	349	6	1
<i>MPI</i>	Mannose phosphate isomerase	Fwd 5'-ACCAGAAATGGCCATTC-3' Rev 5'-GCAGCCATTCACATTAAT-3'	486	TTTAAACC	GGGAAGCA	375	11	5
<i>ZWF1</i>	Glucose-6-phosphate dehydrogenase	Fwd 5'-GTTTCATTTGATCTCGAAGC-3' Rev 5'-GCCATTGATAAGTACCTGGAT-3'	702	AAAAACCAGG	TTGAATTAC	491	8	2
<i>CaVPS13</i>	Vacuolar protein sorting protein	Fwd 5'-TCGTTGAGAGATAATCGACTT-3' Rev 5'-ACGGATGGATCTCCAGTCC-3'	741	CCTTGATATG	CAAAATCATGG	403	16	8
<i>CaADP1</i>	ATP-dependent permease	Fwd 5'-TTGATCAACAACCCGATAAT-3' Rev 5'-GAGCCAAAGTATGAATGATTTG-3'	537	CACGTTGCAA	GGAAATCCAA	443	15	8
<i>CaSYA1</i>	Alanyl-RNA synthetase	Fwd 5'-AGAAGAATGTTGCTGTTACTG-3' Rev 5'-GTTACCTTACCACCAGCTTT-3'	543	TAAATCTAAA	AGCTGTATCA	391	13	4
<i>CaRPN2</i>	26S proteasome regulatory subunit	Fwd 5'-TTCATGCTGCTGGTACTAC-3' Rev 5'-TAATCCCATACCCAAAGCAG-3'	447	TTGGTCTAAA	GTCCTTACGA	306	11	2

the most discriminatory genes described by Bougnoux and colleagues (2) and of four other sequences chosen on the basis that the encoded enzymes were shown previously to be polymorphic in multilocus enzyme electrophoresis analyses (26). Our results allow us to propose an improved gene set for *C. albicans* MLST, and we demonstrated that MLST data for *C. albicans* showing two bases at a single site indeed represent allelic heterozygosity and not sequencing errors. Finally, we included two non-MLST, DNA-based typing characters in our typing scheme. The first is the subdivision of isolate types as described by McCullough et al., based on sequence variation at the transcribed spacer locus of the 25S rRNA gene (22), which divides the species into three major subtypes of epidemiological significance (21). The second is the determination of homozygosity or heterozygosity at the mating type-like (MTL) locus, originally described by Hull and Johnson (13), which is emerging as being associated with important properties such as antifungal resistance (27) and rapid phenotypic switching (17). We propose a unified scheme for gene sequence-based strain typing of *C. albicans* that is portable, reproducible, and discriminatory.

#### MATERIALS AND METHODS

**Isolates.** The 75 *C. albicans* isolates were from our collection of pathogenic fungi. Sixteen were sequential oral isolates from a human immunodeficiency virus-positive patient and have been described previously (35). Eleven were sequential isolates from different anatomical sites of a patient undergoing chemotherapy for hematological malignancy (24). Each of these two sets of isolates was assumed to be epidemiologically related. Two of the isolates were SC5314, a strain widely used for molecular genetic studies, and CAF2, derived from SC5314 by specific disruption of one copy of the *URA3* gene. This pair of strains was included as a control to check for complete identity of sequences in strains that should not differ by MLST. Strain T26 was derived from SC5314 by a complex lineage that included a spontaneous mutation step (6) and was expected to be either identical or very similar to SC5314.

The remaining 45 isolates were chosen to represent different degrees of genetic and phenotypic diversity based on their date, anatomical site, and geographical source of isolation. The set included 13 vaginal isolates originating from the United States in 1998, intended to represent isolates with a constant anatomical location. The set of 45 isolates plus SC5314, CAF2, T26, and one representative each from the two sequential series constituted our main set of 50 notionally diverse strain types. The yeasts were maintained on Sabouraud agar (Oxoid, Basingstoke, United Kingdom).

**Choice of loci for MLST.** Initially, 10 gene fragments were chosen: five that gave the greatest MLST discrimination in the hands of Bougnoux et al. (2) and five that corresponded to a subset of the 13 *C. albicans* housekeeping genes that were previously shown to be polymorphic in multilocus enzyme electrophoresis experiments (26). We reduced the set to a total of eight gene fragments that gave good discrimination in pilot experiments with 20 isolates of *C. albicans* (Table 1). Use of these two sets of four gene fragments allowed us to make direct comparisons of the results obtained from our full panel of isolates with those for the same four gene fragments already published (2). Primers were designed to amplify gene fragments of 450 to 750 bp and are also detailed in Table 1. The primers 5'-ACTCAAGCTAGATTTTGGC-3' (forward) and 5'-CAGCAACA TGATTAGCCC-3' (reverse), which are specific for the *AAT1a* region of the *AAT1* gene upstream of the conserved region, were used for experiments to investigate heterozygosity in MLST.

**DNA extraction.** Genomic DNA was extracted from yeasts grown in YPD broth, containing 2% glucose, 2% mycological peptone (Oxoid), and 1% yeast extract (Difco, Detroit, Mich.). Briefly, cells were harvested in stationary phase and lysed by vortexing the pellet for 3 min with 0.3 g of glass beads (0.45 to 0.52 mm in diameter; Sigma, St. Louis, Mo.) in 200  $\mu$ l of buffer (100 mM Tris-HCl [pH 8.0] containing 2% Triton X-100, 1% sodium dodecyl sulfate, and 1 mM EDTA) and 200  $\mu$ l of 1:1 (vol/vol) phenol-chloroform solution. After vortexing, 200  $\mu$ l of TE (1 mM EDTA, 10 mM Tris-HCl, pH 8.0) was added to the lysate; the mixture was microcentrifuged at full speed for 10 min, and the aqueous phase was transferred to a new tube. DNA was precipitated by addition of 1 ml of

ethanol to the supernatant. Samples were centrifuged, and the pellet was resuspended in 400  $\mu$ l of TE containing 10  $\mu$ l of 10-mg/ml RNase (Sigma, St. Louis, Mo.). The mixture was incubated for 1 h at 37°C, and then DNA was precipitated with 2 volumes of isopropanol and 10  $\mu$ l of 4 M ammonium acetate, dried, and redissolved in 50  $\mu$ l of TE, pH 8.0.

**Amplification and nucleotide sequence determination.** PCR assays were used to amplify the gene fragments. Reaction volumes of 50  $\mu$ l contained 100 ng of genomic DNA, 2.5 U of *Pfu* DNA polymerase (Promega, Madison, Wis.), 5  $\mu$ l of 10 $\times$  buffer (supplied with the enzyme), 200  $\mu$ M deoxynucleoside triphosphate mix (Promega) and 10  $\mu$ M each of the forward and reverse primers. A Flexigene thermocycler (Techne, Cambridge, United Kingdom) was set up with a first cycle of denaturation for 2 min at 94°C, followed by 25 cycles of denaturation at 94°C for 1 min, annealing at 52°C for 1 min, elongation at 72°C for 1 min, and a final extension step of 10 min at 72°C. The amplified products were purified with a commercial PCR purification system (Wizard PCR preps DNA purification system; Promega, Southampton, United Kingdom). Both strands of purified gene fragments were sequenced on an ABI (Foster City, Iowa) 3700 DNA analyzer with a 2.5  $\mu$ M concentration of the same primers that were used in the PCR step. The sequence data were coupled with DNASTar software. Heterozygosities were defined by the presence of two coincident, equivalently sized peaks in the forward and reverse sequence chromatograms. The one-letter code for nucleotides from the International Union of Pure and Applied Chemistry nomenclature was used to define the results.

**Statistical analysis of MLST data.** To determine similarities between MLST strain types, the nucleotides at all 87 polymorphic loci found for the eight gene fragments and 50 notionally diverse isolates were scored for each pair of isolates as 0 for identical nucleotides, 0.5 for heterozygous or homozygous pairs that shared one nucleotide, and 0.0 for identical nucleotides. A similarity matrix was generated by adding the 87 scores for each pair of isolates and dividing the result by 87. Each pair of isolates was therefore assessed by a similarity index between 0 (complete nonidentity) and 1.0 (complete identity). To represent the matrix in two dimensions, a single-linkage dendrogram was constructed by the unweighted pair group method with arithmetic mean with the aid of Mega software (<http://www.megasoftware.net/>). The same software was used to construct a neighbor-joining tree.

To determine closely related genotypes within highly similar clusters, the data were subjected to a Burst analysis (<http://www.mlst.net/Burst/burst.htm>) that was devised originally for analysis of bacterial MLST data (9, 20). The software determines clonal complexes from the data, suggesting a consensus "ancestral" type which contains the most-represented identical loci for a subgroup of isolates and indicates variants differing at just one, two, or three loci. The results are displayed as concentric circles, with the consensus strain type in the center and each new circle indicating isolates that differ from the consensus sequence by one nucleotide for each circle. The default design of the software analyzes seven gene fragments for five groups. We set our data input to accommodate eight gene fragments and scrutinized the results with group settings of three, four, five, and six. The five-group and six-group settings gave identical and interpretable results.

The discriminatory power (*D*) of the MLST system was determined by the formula of Hunter (14).

**Investigation of heterozygous loci.** The accuracy of determination of heterozygosity was ascertained with experiments done with the *AAT1a* fragment (478 bp) from *C. albicans* 76/002, which showed four putatively heterozygous sites. It was cloned in the pGEM-T Easy vector system (Promega). Because *Pfu* polymerase generates blunt ends, the PCR product was A-tailed by incubating 5  $\mu$ l of the purified gene fragment at 70°C for 30 min in the presence of 1  $\mu$ l of 10 $\times$  buffer, 1  $\mu$ l of 25 mM MgCl<sub>2</sub>, 1  $\mu$ l of 10 mM dATP, and 5 U of *Taq* polymerase. The DNA ligation reaction mix comprised a 10- $\mu$ l volume containing 1  $\mu$ l of the A-tailed PCR product, 1  $\mu$ l of 50-ng/ $\mu$ l pGEM-T vector, 5  $\mu$ l of 2 $\times$  rapid ligation buffer (supplied with the enzyme), and 3 U of T4 DNA ligase (Promega), and the mixture was incubated at 4°C overnight. Then 5  $\mu$ l of ligation mix was used to transform *Escherichia coli* XL-1-Blue competent cells, following the method described by Hanahan (12). Transformed cells were subsequently plated on Luria-Bertani plates with ampicillin, 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside, and isopropylthiogalactopyranoside (IPTG) (12).

Plasmid DNA was extracted from six colonies of *E. coli* that grew on this medium with the Qiaprep Spin Miniprep kit (Qiagen, West Sussex, United Kingdom), following the manufacturer's protocol. The presence of the expected gene fragments was separately checked by digesting the plasmid DNA obtained from the six selected clones with *Eco*RI (New England Biolabs, Beverly, Mass.) and by PCR amplifying the *AAT1* fragment from plasmid DNA. Plasmid DNA obtained from the six clones was then sequenced as previously described, with a 5  $\mu$ M concentration of the same primers that were used in the PCRs.

Second, we used *Msp*I to digest the *AAT1a* fragment from *C. albicans* 76/002.

Sequence analyses had shown that the heterozygosities in this PCR product should result in the creation of an *Msp*I restriction site in one of the alleles at polymorphic site 6. No *Msp*I restriction site was found anywhere else in the whole gene fragment. The gene fragment was digested for 4 h at 37°C with *Msp*I (New England Biolabs) in a 30- $\mu$ l reaction volume containing 5  $\mu$ l of the PCR product, 3  $\mu$ l of 10 $\times$  buffer 2 (supplied with the enzyme), and 1.5  $\mu$ l of 20-U/ $\mu$ l *Msp*I. Digestion products were loaded onto a 1.8% agarose gel containing ethidium bromide (0.5  $\mu$ g/ml). TAE (40 mM Tris acetate [pH 8.0], 1 mM EDTA) was used as the running buffer, and a 100-bp DNA ladder (Promega) was used as molecular size markers. DNA bands were visualized by UV transillumination.

**Additional strain typing characters.** PCR for MTL status used the primers Fwd (5'-GAATTCACATCTGGAGGC-3') and Rev (5'-CAAAGCAGCCAAC TCAGG-3') for *MLT $\alpha$*  and Fwd (5'-ACCTGCATGAAGAAACAG-3') and Rev (5'-GTGGCTAGGTTGAATTTG-3') for *MLLa*. Conditions were as described above, but 50- $\mu$ l multi-PCR volumes contained 100 ng of genomic DNA, 2.5 U of *Taq* polymerase (Promega), 5  $\mu$ l of 10 $\times$  magnesium-free buffer, 3  $\mu$ l of 25 mM MgCl<sub>2</sub>, 200  $\mu$ M deoxynucleoside triphosphate mix, and 5  $\mu$ M each of the forward and reverse primers. PCR for the rRNA gene transcribed spacer region was done as previously described (22).

## RESULTS

**Optimization of MLST.** A total of eight amplicons based on four previously used sequences and four new sequences were used for all MLST analyses (Table 1). For the four new gene fragments (*AAT1a* to *ZWF1* in Table 1), sequences ranging from 339 to 500 bp were obtained from nominal amplicon sizes of 478 to 702 bp. For the four gene fragments described previously (2), we obtained sequences of comparable or greater length. Across all eight gene fragments sequenced, 6 to 16 polymorphic loci were found among our 75 test isolates. The sequence differences equated to one to seven polymorphic amino acid sites per gene fragment.

For the previously studied gene fragment derived from *CaVPS13*, we found a further four polymorphic sites upstream and one further site downstream of the portion of sequence already published (2). For *CaSYA1*, an additional two polymorphic sites upstream of the published sequence were found, and for *CaRPN2* one more polymorphic site was revealed downstream of the published sequence. The data for polymorphic nucleotide sites in Table 1 for the four gene fragments already published are limited to the range of the published sequences and do not include these extra polymorphisms. The results show that the isolates that were investigated revealed two new polymorphic sites (positions 157 and 350) within the published range for *CaADP1* and two (positions 32 and 307) for *CaSYA1*.

**Nucleotide polymorphisms and amino acid changes.** To investigate the impact of nucleotide polymorphisms on amino acid sequence, we mapped the triplet codons for each gene fragment, based on the genomic information available (from the Stanford [<http://www-sequence.stanford.edu/group/candida/>], Galar Fungail [[http://www.pasteur.fr/recherche/unites/Galar\\_Fungail/](http://www.pasteur.fr/recherche/unites/Galar_Fungail/)], and Minneapolis [<http://alces.med.umn.edu/bin/genelist?genes>]) for the *C. albicans* genome databases. While most of the polymorphisms were synonymous, 35 (40%) of the 87 individual changes recorded were nonsynonymous. Details of the alterations are shown in Table 2. Of the 35 amino acid changes, 19 were substantive changes, e.g., basic to acidic side chains, aliphatic to aromatic side chains, etc. In two cases, the change was between proline and serine.

The Appendix details the polymorphisms for the eight MLST gene fragments that were used in this study and the



TABLE 2. Changes in amino acid sequence resulting from nucleotide polymorphisms

Fragment	Starting nucleotide position of triplet	Triplet polymorphism	Amino acids
<i>AAT1a</i>	20	C/T C A	P/S
<i>AAT1b</i>	31	C/T C T	P/S
<i>MPI</i>	34	G/A C T/C	A/T
	70	G A A/T	D/E
	88	A/G T C	I/A
	232	G A C/A	D/E
	289	A/G G A	R/G
<i>ZWF1</i>	23	A/C A A	K/Q
	482	A/T T T	I/F
<i>CaVPS13</i>	32	C G/A A	R/Q
	134	A/T G T/C	S/C
	212	G/A T T	V/I
	281	A/G A/G A	R/G/K/E
	321	G/A T G/T	V/M/I
	326	G/A T G/A	V/M/I
	368	A A T/A	N/K
	374	T T/A C	F/Y
	41	C A/G C	H/R
<i>CaRPN2</i>	200	G A C/A	D/E
	35	A/G A A	K/E
<i>CaADP1</i>	125	A/G T T	I/V
	157	A A/G T	S/N
	215	G/A T C	V/I
	224	C T/A A	L/Q
	263	A/C A A	N/Q
	281	G G/C T	G/A
	395	A/T A C	N/Y
	32	A/G A A/C	K/N/D/E
<i>CaSYA1</i>	185	G/C C T	A/S
	323	G A/G A	E/G
	350	G C/T C	A/V

diversity of genotypes found based on these sequences (Table 1A). At some polymorphic loci, a majority of the 50 notionally diverse isolates had the same sequence, while others showed more interisolate diversity. The genotype numbers assigned in the previous study were used for *CaADP1*, *CaRPN2*, *CaSYA1*, and *CaVPS13*. Many new genotypes were determined in our set of isolates (Appendix) which had not been reported (2). Genotypes represented by large numbers of our isolates for each of the gene fragments shown in the Appendix were also represented by many isolates in the previous study. For example, for *CaADP1*, 9 of our 50 isolates gave new strain types (numbered from 17 upwards), for *CaRPN2* we found four new types (17 to 20), represented by five isolates, for *CaSYA1* there were 14 new types (14 and 27), each represented by a single isolate, and for *CaCAVPS1* there were 20 new types (25 to 44), represented by 26 isolates. The MLST genotypes found in this study are summarized in Table 3. The results show a discriminatory power of  $D = 0.996$  (14).

**Heterozygosity and MLST genotypes.** The data in the Appendix, like the equivalent data already published (2), show that sequence heterozygosity contributed considerably to the diversity of genotypes determined by MLST. For our panel of 50 *C. albicans* isolates (Table 3), a total of 87 polymorphic nucleotide sites were found across all eight gene fragments tested. For only 16 (18.4%) of these sites did the polymorphisms consist entirely of homozygous nucleotide changes; the rest always included at least one example of heterozygosity at

the site. For each polymorphic site, only three sequence results were obtained: one of two bases or the heterozygous combination of the same two bases. No polymorphic site resulted from more than two nucleotide changes. The heterozygous PCR products obtained from the diploid genome of *C. albicans* were presumed to arise from the coamplification of both alleles, resulting in sequence profiles that showed two coincident peaks in the sequence chromatogram (Fig. 1). However, the high prevalence of apparent heterozygosity and the occasional ambiguity in double peaks (Fig. 1a) raised the possibility that sequencing errors rather than true heterozygosities generated patterns of this type.

**Cloning and sequencing of *AAT1a* gene fragment.** To investigate further whether double peaks of the type shown in Fig. 1 represented true allelic heterozygosities or sequencing errors, we chose the *AAT1a* gene fragment because sequence analysis of the PCR product showed that the presumed coamplification of two different bases could create a unique *MspI* restriction site in one of the alleles.

*C. albicans* isolate 76/002 had four potentially heterozygous sites in the *AAT1a* sequence (genotype 1 in the Appendix). At position 40, the computer analysis of the sequence showed two equally sized peaks for adenine and guanine only for the reverse direction (Fig. 1a), while the putative heterozygosity was detected by sequencing in both directions at nucleotide 124 (Fig. 1b). The putative heterozygosities at loci 7 and 89 were of the clear double peak variety shown in Fig. 1b. Six colonies of *E. coli* transformed with the cloned *AAT1a* PCR fragment were selected randomly for plasmid DNA extraction and sequencing analysis. The results obtained showed that four of the six clones, each theoretically containing one of the two alleles of isolate 76/002, carried the bases G, A, A, and C at the four polymorphic sites, while for the other two clones the bases A, G, G, and T were identified in the same nucleotide positions.

**Validation of *MspI* polymorphism.** One of the putative sequence heterozygosities (position 124, Y = C or T) observed in the PCR product in *C. albicans* 76/002 created an *MspI* restriction site (C<sup>+</sup>CGG) in one of the alleles. Since no *MspI* restriction sites were found anywhere else in the gene fragment, the PCR product was digested with *MspI*. As shown in Fig. 2, the digested products confirmed that the allele with the *MspI* restriction site gave the two predicted DNA fragments of 305 and 173 bp, while the other one remained undigested, as evidenced by the DNA band of 478 bp. Therefore, sequencing errors were unlikely to account for the polymorphisms observed in MLST for this diploid organism.

**Additional strain typing characters.** Among our 50 nominally distinct isolates, four were homozygous at the MTL locus: J990578, 85/005, and 81/225 were type  $\alpha$ , while S9 was type  $\alpha$ . The majority of the isolates (40 of 50) were genotype A by the transcribed spacer element PCR. There were seven type B strains and three type C strains. These characters are shown, together with the MLST data, in Table 3.

**Epidemiological relationships of isolates typed.** In Table 3 the diversity of genotypes detected across all eight MLST fragments is shown for all 50 isolates tested. For the four previously studied gene fragments (2), the published genotype numbers are used, and additional, higher numbers are assigned for the novel genotypes that we found in this study (Table 3). The final column indicating the full range of diploid sequence ge-

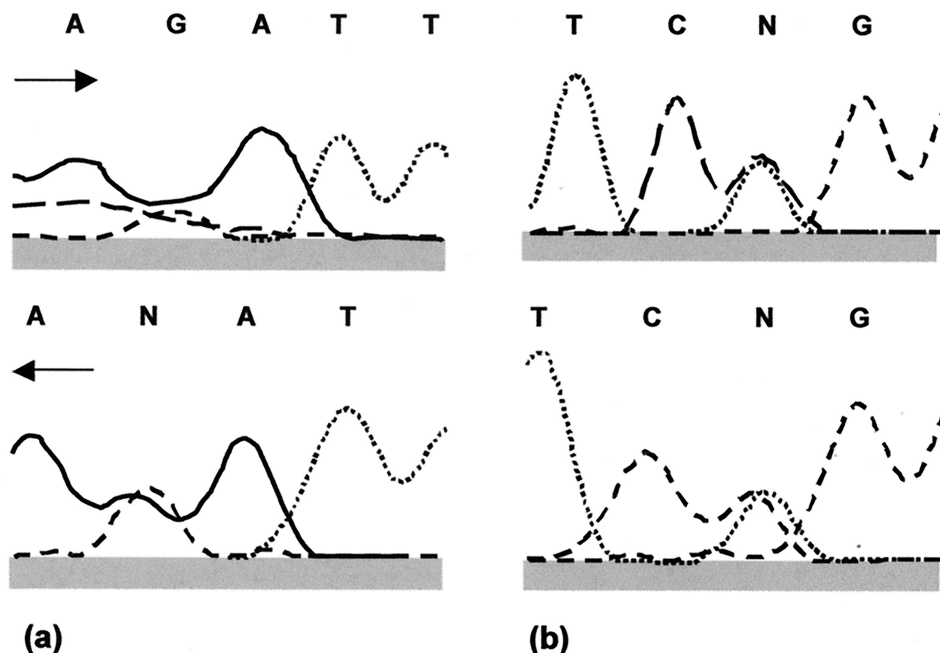


FIG. 1. Two examples of raw sequencing data for forward and reverse strands of PCR product from *C. albicans* strain 76/002, *AAT1a* gene fragment, illustrating sequence results interpreted as heterozygosity. Solid line, adenine (A); long-dashed line, cytosine (C); short-dashed line, guanine (G); dotted line, thymine (T); N, site of putative heterozygosity because the peaks suggest that two nucleotide bases are present in the sequence. (a) Nucleotide 40; (b) nucleotide 124. Arrows indicate direction of sequencing.

notypes (DSTs: the combination of genotypes from all individual gene fragments) indicates that 46 unique types were found among the 50 isolates examined. Two identical DSTs were found for SC5314 and its derivative CAF2. An oral isolate, 78/028, from a healthy volunteer first cultured in 1978, and J981305, from a patient with vaginitis in the United States, were found to be identical to type 22. Three isolates, two from different U.S. patients with vaginitis and an isolate from a penis

obtained 17 years earlier in the United Kingdom, shared DST 28.

A UPGMA similarity dendrogram from the data for the 50 isolates tested was generated with Mega software (Fig. 3). The two pairs and triad of isolates with identical MLST types, together with a further 19 isolates, clustered in a single group (bracketed in Fig. 3) with >84% identity by this method of analysis. The remaining isolates generally showed a higher level of diversity. Analysis of the strain types by means of a neighbor-joining tree similarly clustered the same 26 isolates bracketed in Fig. 3 into a single, highly related group. Nine of the 13 vaginal isolates obtained from the United States in 1998 fell into the highly similar clusters in both analyses, with the remainder showing little relation to each other.

Although single-linkage cluster analysis of the type shown in Fig. 3 permits sorting of isolates into similar clusters, MLST data allow a more refined analysis of isolates based on the relationships between very closely related genotypes within clonal complexes. This model postulates a putative ancestral genotype from which other types have developed by mutation at just one or two loci (9, 20). Analysis with the Burst algorithm revealed three such clonal complexes for our isolates; however, two of the putative clonal complexes included just two or three isolates. The largest complex, which was therefore capable of structural analysis, is shown in Fig. 4. The complex was in fact composed of two related subcomplexes, which divided our main UPGMA cluster (Fig. 3) into three sets. The first comprised isolates J981303, J981309, 81/190, 78/028, J981305, 83/004, 76/002, 81/133, J981301, 81/193, B59630, J981314, and 81/192; the second comprised 2-76, J981307, 85/045, T26,

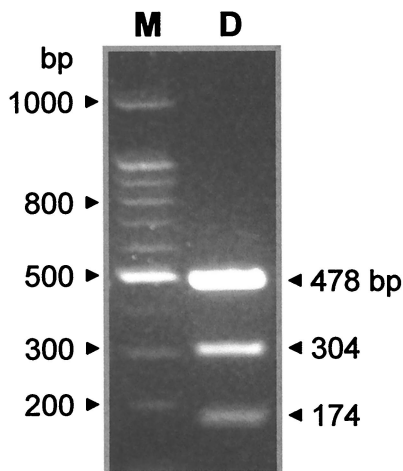


FIG. 2. *MspI* restriction digestion of *AAT1a* PCR product from *C. albicans* strain 76/002. Electrophoresis of the digest revealed three bands (lane D), indicating the products predicted when two heterozygous sequences are treated with the enzyme and only one contains the restriction site. Lane M, 100-bp ladder.

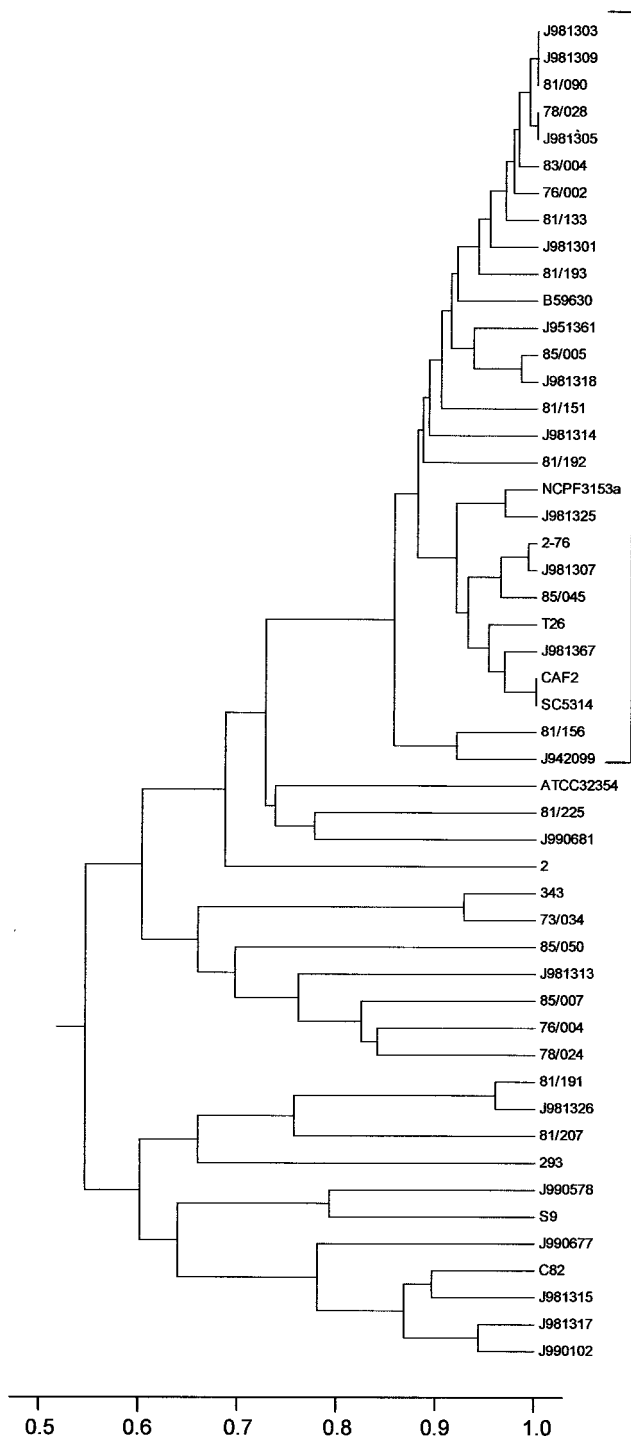


FIG. 3. Single-linkage dendrogram indicating the similarities of 50 *C. albicans* isolates determined by MLST with eight gene fragments.

CAF2, and SC5314; and the third comprised the remainder of the isolates from the related cluster.

**MLST patterns of sequential isolates from the same patients.** The 16 sequential isolates from an AIDS patient (35) all gave the same diploid sequence type (DST) by MLST. Of the 11 consecutive isolates from oral and fecal surveillance cul-

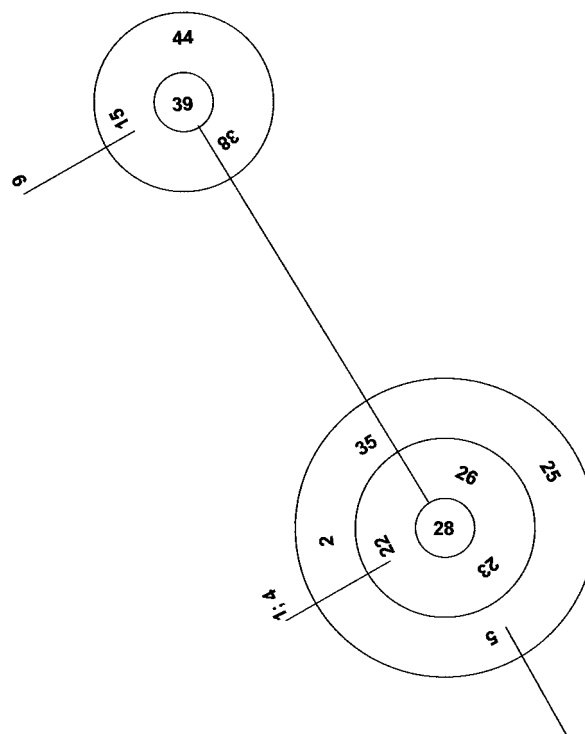


FIG. 4. Burst analysis showing 16 clonally related diploid sequence type numbers (see Table 3) derived from a subset of the isolates bracketed in Fig. 3. In each set of concentric circles, the central type has a common consensus sequence between isolates for the eight gene fragments tested; each surrounding circle indicates one level of change.

tures from a single patient undergoing chemotherapy for hematological malignancy, nine gave one DST 40 and two gave a DST that differed at a single polymorphic site in *CaADP1*, which was heterozygous (A or G) in the nine isolates and homozygous (G) at this locus in two, suggesting that MLST was sufficiently sensitive to detect microsequence evolution within the clade.

**DISCUSSION**

The set of gene fragments we had chosen for MLST, based on multilocus enzyme electrophoresis data (26), differed from the set published by Bounoux et al. (2), and the appearance of their publication therefore gave us the opportunity to validate the published data and to add our own MLST gene fragment set to facilitate the selection of an optimal set of gene fragments that could be used routinely for *C. albicans* MLST. The results of this study confirm that both the published set of gene fragments and those we chose were able to indicate genotypic diversity among isolates. The set of eight gene fragments listed in Table 1 all give high genotypic diversity with nonoverlapping results: 49 nonidentical isolates (SC5314 and CAF2 were expected to be identical) gave 46 different diploid sequence types on the basis of these sequences (Table 3). This high level of differentiation reinforces the view of MLST as a highly discriminatory strain-typing procedure.

Data for highly related isolates show that MLST gives high

TABLE 3. Genotypes of 50 *C. albicans* isolates analyzed by MLST

Strain	Site of isolation or source	Origin	Genotype								DST type	ABC type <sup>a</sup>	MLT type <sup>b</sup>
			AAT1a	AAT1b	MPI	ZWF1	CaADP1	CaSYA1	CaRPN2	CaVPS13			
B59630	Oral (AIDS)	Germany	1	1	1	1	12	4	14	25	1	A	Het
81/193	Sputum	USA	1	1	1	2	13	14	14	21	2	A	Het
85/007	Wound	London, UK	2	2	2	3	17	15	17	26	3	A	Het
76/002	Mouth	Leicester, UK	1	1	1	4	12	4	14	25	4	A	Het
83/004	Mouth	Germany	1	1	3	2	13	4	14	27	5	A	Het
J951361	Blood	Glasgow, UK	3	1	3	2	13	4	14	28	6	A	Het
J990102	Vagina	Belgium	4	3	4	5	10	16	13	29	7	B	Het
J990578	Mouth	France	5	3	5	6	18	17	18	30	8	A	a
T26	Derived from SC5314		3	4	6	2	13	4	14	3	9	A	Het
2	Cutaneous nodule	Spain	6	5	7	7	19	18	14	30	10	C	Het
S9	Sputum	Italy	4	6	8	6	18	19	3	31	11	C	α
J942099	Unknown	USA	1	7	2	8	13	4	14	7	12	A	Het
J981367	Vagina	USA	7	1	6	8	13	4	14	32	13	A	Het
C82	Mouth (AIDS)	Switzerland	4	3	9	9	20	20	13	33	14	B	Het
CAF2	Derived from SC5314		1	1	6	2	13	4	14	3	15	A	Het
SC5314	[Laboratory isolate]		1	1	6	2	13	4	14	3	15	A	Het
293	Nose	London, UK	8	8	10	1	21	21	3	34	16	A	Het
NCPF 3153a	Reference isolate	UK	3	1	10	2	12	4	14	35	17	A	Het
ATCC 32354	=B311, reference isolate	USA	9	9	11	1	1	22	14	36	18	A	Het
J981315	Vagina	USA	4	3	4	5	14	9	14	33	19	B	Het
73/034	Mouth	Leeds, UK	3	2	2	10	15	9	4	11	20	B	Het
J990681	Blood	Canada	10	10	1	3	1	23	3	37	21	A	Het
78/028	Mouth	Leicester, UK	1	1	1	2	12	4	14	7	22	A	Het
J981305	Vagina	USA	1	1	1	2	12	4	14	7	22	A	Het
J981301	Vagina	USA	1	1	1	2	13	4	14	3	23	A	Het
J990677	Blood	Canada	4	6	1	5	10	24	14	38	24	B	Het
81/192	Unknown	USA	2	1	1	2	13	4	14	21	25	A	Het
81/133	Feces	USA	1	1	1	11	13	4	14	7	26	A	Het
81/207	Blood	USA	10	3	6	12	22	25	19	30	27	A	Het
81/090	Penis	Leicester, UK	1	1	1	2	13	4	14	7	28	A	Het
J981303	Vagina	USA	1	1	1	2	13	4	14	7	28	A	Het
J981309	Vagina	USA	1	1	1	2	13	4	14	7	28	A	Het
76/004	Mouth	Leicester, UK	2	2	2	13	1	8	15	5	29	A	Het
J981313	Vagina	USA	11	2	12	14	1	8	14	39	30	A	Het
81/151	Mouth	USA	12	11	1	15	1	4	14	27	31	A	Het
J981326	Vagina	USA	13	3	13	16	6	12	20	40	32	A	Het
81/191	Mouth	USA	14	3	13	16	22	12	20	41	33	A	Het
J981317	Vagina	USA	4	3	4	17	10	9	13	33	34	B	Het
J981314	Vagina	USA	1	1	1	2	12	8	14	7	35	A	Het
J981325	Vagina	USA	3	1	6	18	12	4	14	7	36	A	Het
78/024	Mouth	Leicester, UK	2	2	1	19	23	26	15	42	37	A	Het
J981307	Vagina	USA	1	1	6	2	1	4	14	7	38	A	Het
2-76 <sup>c</sup>	Mouth (AIDS)	USA	1	1	6	2	13	4	14	7	39	A	Het
343 <sup>d</sup>	Mouth	London, UK	3	2	2	22	1	9	3	44	40	B	Het
81/156	Faeces	USA	1	1	1	3	1	4	14	6	41	A	Het
85/050	Vagina	Leicester, UK	15	2	1	21	16	10	9	43	42	C	Het
85/005	Blood	London, UK	3	4	1	15	1	4	14	7	43	A	a
85/045	Anus	Leicester, UK	1	1	6	20	13	4	14	7	44	A	Het
81/225	Blood	USA	10	9	13	20	1	27	14	7	45	A	a
J981318	Vagina	USA	3	4	1	2	1	4	14	6	46	A	Het

<sup>a</sup> Determined by PCR of transcribed spacer region of rDNA region.  
<sup>b</sup> Determined by multiplex PCR of MLT locus.  
<sup>c</sup> From a series of 16 sequential isolates (35).  
<sup>d</sup> From a series of 12 isolates from a single patient with a hematological malignancy.

reproducibility within and between laboratories. Isolate SC5314 was also tested by MLST in the study by Bougnoux and colleagues; the SC5314 genotypes found by them (2) were identical to those we determined with the same DNA fragments. CAF2, derived from SC5314 by deletion of one copy of the *URA3* gene, also gave the same result. This consistency is a demonstration of the power and reproducibility of MLST applied to *C. albicans*. Unlike some typing systems, in which reproducibility and discriminatory power are inversely related (14), our data show both 100% reproducibility and a discriminatory power of 0.996 (14). Moreover, sequential isolates from each of two patients gave identical MLST types in one instance and types that differed by only a single nucleotide in the second case. These findings demonstrate that MLST with the gene

fragments used in this study can recognize isolates of *C. albicans* that are identical or nearly so and may be detecting minor microevolutionary changes, known to occur in longitudinal studies with repeated isolates from the same patient (16, 25, 31).

That such changes may occur even in laboratory isolates is exemplified by the finding of four nucleotide differences between strain T26 and its parents SC5314 and CAF2. T26 was engineered from SC5314 by a series of changes that included gene disruptions and selection for spontaneous resistance to echinocandins (6, 15). We conclude that one or more of the steps in the lineage of T26 resulted in the small sequence differences detected in this study. All four nucleotide changes in strain T26 occurred in a single strand of the diploid DNA in



the *AATI* gene, since each involved loss of nucleotide heterozygosity.

Among the amino acid changes resulting from 40% of the nucleotide polymorphisms, many involved switches between types of amino acid, including two instances of proline to serine (Table 2), which would be expected to effect significant alterations in secondary and higher peptide structures. If this level of sequence change is representative of variation for the products of other *C. albicans* genes, it must be concluded that the fungus is tolerant of the observed differences in protein structure. It is likely that many subtle phenotypic differences between strains may exist that have yet to be detected. The high level of genotype-phenotype differences possible between strains of a yeast species was indicated in a recent study based on expression profiling with a fresh, wild-type isolate of *Saccharomyces cerevisiae* and a laboratory-maintained isolate. This analysis showed that 1,500 of the 6,116 genes in the yeasts differed in levels of expression (3). To what extent the differences in expression are reflected as differences in the genome sequence have yet to be determined.

The findings of this study indicate that the set of genes listed in Table 1 are adequate for high-quality strain typing by MLST. For population genetic analyses and other statistical approaches to the epidemiological study of *C. albicans*, this larger set of MLST fragments should represent a more discriminatory tool than the six-fragment set already proposed (2). It is notable that the sequence differences between SC5314 and T26 would not have been detected with the published six-fragment set. These two strains are not the only examples of isolates whose types were indistinguishable by MLST with the four published gene fragments but could be distinguished by the new gene fragments. Conversely, some strains could be distinguished by the published gene fragments but not by the new ones.

So far, *C. albicans* is the only example for which MLST has been attempted with a species having a permanently diploid genome, where heterozygous alleles may occur. The frequency with which heterozygous sites occur in MLST with *C. albicans* is high, both in our own and in the previous study of MLST (2). The present study investigated the possibility that apparent heterozygosity may arise through sequencing artifacts and confirmed it to be genuine. Heterozygosity in a diploid genome adds extra characters for strain discrimination over simple sequence variation and will allow future analyses of population genetics research based on haplotypes. The relative frequencies of clonal reproduction and sexual or other recombinational events in natural populations of the fungus remain an open question (1, 10, 11, 16, 18, 26, 30, 34).

The addition of two non-MLST characters to a *C. albicans* strain-typing scheme adds extra discriminatory detail. Data from surveys based on the system that divides *C. albicans* into three subtypes, A, B, and C, based on sequence differences in the transcribed spacer region in the gene encoding rRNA have already been used to demonstrate geographical differences in *C. albicans* isolate populations (21). This approach also allows direct, unequivocal recognition of the species *C. dubliniensis* (22). In common with McCullough et al. (21), we found most of our isolates were type A by rDNA PCR. Of note, all the isolates in the cluster of highly related strains (Fig. 3) were type

A. The level of discrimination possible by MLST clearly exceeds that of ribosomal DNA typing, since we could distinguish 36 distinct strains among the 40 designated as type A. All the isolates of type B and type C strains could be differentiated by MLST.

Determination of mating type in *C. albicans* isolates is of possible relevance to antifungal resistance (27) and to phenotypic switching in this fungus (17). In common with Rustad and colleagues (27), we found that only a minority of isolates were homozygous at *MTL*. They found 12 homozygous isolates among 96 tested (12.5%); we found 4 among 50 (8%). Even if the frequency of homozygous mating types in the clinical population of *C. albicans* is only on the order of 10%, this may be adequate to explain small departures from clonality in the Hardy-Weinberg equilibrium analyses that have been described previously for this species (11, 26).

We conclude that MLST with *C. albicans* offers an effective system for epidemiological work with the species, that the high frequency of heterozygous sequences in the DNA regions chosen for MLST add extra information to MLST that is not available with haploid organisms, and that the creation of a central database for archiving of MLST data will enhance research based on strain typing. Although at present MLST is more likely to constitute a research rather than a reference tool, MLST has the advantage that it is scalable from a small number of isolates to many hundreds or even thousands of isolates by the exploitation of robotic DNA extraction and high-throughput nucleotide sequence determination technologies. The application of high-throughput technology also leads to substantial reductions in the cost of isolate characterization. For example, in this study full MLST profiles were obtained for a consumables cost of approximately US\$30 per isolate; however, with recently developed DNA analyzers, a cost of less than US\$15 per isolate is now attainable, with the prospect of further substantial reductions in cost in the near future. Such automation will not only reduce costs but also increase throughput of the method. At present, at least 24 isolates per week can be typed with all procedures done by hand, but automation of the processes will increase this number to hundreds per week.

Inclusion of the *MTL* status and genotype (A, B, or C) as extra typing data further refines the ability of DNA-based methods to distinguish isolates of *C. albicans*. The findings of this study and the previous investigation (2) show a high level of sequence variation in transcribed housekeeping genes among isolates of *C. albicans*. We are now establishing MLST for other *Candida* species and investigating the frequencies of sequence changes in isogenic strains of *C. albicans* exposed to various conditions in vitro and in vivo.

## APPENDIX

The positions of polymorphic nucleotide sites identified in various gene fragments are shown in Table. A1. All the nucleotides present at each variable site are shown for genotype 1. For the other genotypes, only nucleotides that differ from those of genotype 1 are shown, and nucleotides identical to those in genotype 1 are indicated with a dot. The number of isolates from the set of 50 diverse isolates with the same genotype is shown in parentheses. The position of each polymorphic site is indicated for each fragment.



TABLE A1. Polymorphic sites

Fragment	Genotype <sup>a</sup> (no. of isolates)	Nucleotide <sup>b</sup> at position:														
		7	20	40	70	89	124	361								
<i>AAT1a</i>		7	20	40	70	89	124	361								
	1 (19)	R	C	R	T	R	Y	C								
	2 (4)	A	.	A	C	A	C	.								
	3 (8)	A	.	G	.	G	T	.								
	4 (6)	A	.	A	.	A	C	.								
	5 (1)	.	.	G	.	A	.	Y								
	6 (1)	A	.	G	.	A	T	T								
	7 (1)	A	.	.	.	.	.	.								
	8 (1)	A	.	G	Y	G	.	.								
	9 (1)	.	Y	A	Y	A	.	.								
	10 (3)	.	.	A	Y	A	.	.								
	11 (1)	A	.	A	Y	A	C	.								
	12 (1)	G	.	A	.	A	C	.								
	13 (1)	.	.	.	.	A	.	Y								
	14 (1)	.	.	.	.	A	.	.								
	15 (1)	A	.	G	.	A	T	.								
<i>AAT1b</i>		31	42	87	189	204	261									
	1 (23)	C	C	A	T	T	R									
	2 (7)	.	A	G	C	C	G									
	3 (8)	.	.	.	Y	.	G									
	4 (3)	.	.	.	.	.	A									
	5 (1)	.	A	.	C	C	G									
	6 (2)	.	.	.	C	.	G									
	7 (1)	Y	.	.	.	.	A									
	8 (1)	.	.	R	.	Y	.									
	9 (2)	.	.	R	Y	Y	G									
	10 (1)	.	M	R	Y	Y	G									
	11 (1)	.	.	.	.	.	G									
<i>CaADPI</i>		35	40	46	109	125	159	166	205	215	225	232	263	282	352	395
	1* (11)	A	C	C	A	A	G	G	G	G	T	C	A	G	A	A
	2* (0)	.	.	.	.	R	.	.	.	.	.	.	M	.	.	.
	3* (0)	.	T	T	G	.	.	A	A	.	.	T	.	.	.	.
	4* (0)	.	T	T	G	G	.	A	A	.	A	T	.	.	.	.
	5* (0)	.	T	T	G	G	.	A	A	.	A	T	.	S	.	W
	6* (1)	.	T	T	G	R	.	A	A	.	A	T	.	.	.	.
	7* (0)	.	Y	Y	G	G	.	A	A	.	W	T	.	.	.	.
	8* (0)	.	Y	Y	R	G	.	R	R	.	.	Y	.	.	.	.
	9* (0)	.	Y	Y	R	G	.	R	R	.	.	T	.	.	.	.
	10* (3)	.	Y	Y	R	G	.	R	R	.	W	Y	.	.	.	.
	11* (0)	.	Y	Y	R	R	.	R	R	.	.	T	.	.	.	.
	12* (7)	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	13* (16)	R	.	.	.	.	.	.	.	.	.	.	.	.	.	.
	14* (1)	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.
	15* (1)	.	.	.	.	.	.	.	.	R	.	.	.	.	.	.
	16* (1)	.	Y	Y	R	.	.	R	R	.	W	Y	.	.	.	.
	17 (1)	.	.	.	.	R	R	.	.	.	.	.	.	.	.	.
	18 (2)	.	Y	Y	R	G	.	R	R	.	W	Y	.	.	.	.
	19 (1)	.	T	T	G	G	.	A	A	.	A	T	.	C	.	T
	20 (1)	.	.	.	.	.	.	R	R	.	W	Y	.	.	.	.
	21 (1)	.	.	.	.	.	.	.	.	.	.	.	.	.	R	.
	22 (2)	.	Y	Y	R	R	.	R	R	.	W	Y	.	.	.	.
	23 (1)	.	.	T	G	R	.	A	A	.	A	T	.	.	.	.
<i>MPI</i>		21	27	34	36	66	72	88	234	237	276	289				
	1 (20)	G	R	R	Y	K	W	R	C	A	G	A				
	2 (5)	.	G	A	C	G	T	G	.	.	.	.				
	3 (2)	.	G	.	.	.	.	.	.	.	.	.				
	4 (3)	.	G	A	C	G	T	G	M	R	R	R				
	5 (1)	R	G	A	C	G	T	G	M	R	R	R				
	6 (9)	.	A	G	T	T	A	A	.	.	.	.				
	7 (1)	.	A	G	T	T	A	G	.	.	.	.				
	8 (1)	.	G	A	C	G	T	G	A	G	A	.				

Continued on following page

TABLE A1—Continued

Fragment	Genotype <sup>a</sup> (no. of isolates)	Nucleotide <sup>b</sup> at position:															
		21	27	34	36	66	72	88	234	237	276	289					
<i>MPI</i>		21	27	34	36	66	72	88	234	237	276	289					
	9 (1)	.	G	A	C	G	T	G	.	R	R	R					
	10 (2)	.	A	G	T	T	A	A	.	G	.	.					
	11 (1)	.	G	A	C	T	.	A	.	.	.	.					
	12 (1)	.	.	.	.	.	.	.	M	R	R	R					
	13 (3)	.	A	G	T	T	A	.	.	.	.	.					
<i>CaRPN2</i>		19	34	42	49	55	76	130	142	193	202	274					
	1* (0)	G	A	A	A	G	G	G	A	T	C	C					
	2* (0)	.	T	.	.	.	.	.	.	.	M	.					
	3* (4)	.	T	.	.	.	.	.	.	.	.	G					
	4* (1)	.	T	.	.	.	.	.	.	.	.	S					
	5* (0)	.	T	.	.	.	.	.	.	.	.	Y					
	6* (0)	.	T	R	.	.	.	R	R	.	.	.					
	7* (0)	.	T	R	.	.	.	.	.	.	.	S					
	8* (0)	K	T	.	R	R	.	A	.	.	.	.					
	9* (1)	K	T	.	R	R	R	.	.	.	.	.					
	10* (0)	K	T	.	R	R	R	.	.	.	.	S					
	11* (0)	K	T	.	R	R	R	.	.	Y	.	S					
	12* (0)	K	T	R	R	R	R	R	R	.	.	.					
	13* (3)	K	W	.	R	R	R	.	.	.	.	.					
	14* (34)	T	T	.	G	A	A	.	.	.	.	.					
	15* (2)	K	T	.	R	R	R	.	.	.	M	.					
	16* (0)	K	W	.	.	R	R	.	.	.	.	.					
	17 (1)	.	T	.	.	.	.	.	.	.	A	.					
	18 (1)	.	T	.	.	.	.	.	.	.	.	.					
	19 (1)	K	T	.	R	R	A	.	.	.	.	.					
	20 (2)	.	T	.	.	.	.	R	.	.	.	S					
<i>CaSYA1</i>		1	25	32	34	37	61	100	142	160	185	307	324	351			
	1* (0)	C	C	A	A	A	G	T	G	C	G	C	G	Y			
	2* (0)	T	M	.	M	M	R	C	A	Y	.	.	.	C			
	3* (0)	T	A	.	.	.	A	C	A	T	T	.	R	C			
	4* (26)	T	A	.	.	.	.	C	A	T	.	.	.	C			
	5* (0)	T	A	.	.	.	.	C	A	T	.	.	.	.			
	6* (0)	T	A	.	.	.	A	.	.	.	.	.	.	T			
	7* (0)	T	M	.	.	.	A	.	.	.	.	.	.	.			
	8* (3)	T	M	.	.	.	A	Y	R	Y	K	.	.	T			
	9* (4)	T	M	.	.	.	R	Y	R	Y	.	.	.	.			
	10* (1)	Y	.	.	.	.	A	Y	R	.	K	.	.	.			
	11* (0)	Y	.	.	.	.	R	.	.	.	.	.	.	T			
	12* (2)	Y	.	.	.	.	R	Y	R	Y	.	.	.	C			
	13* (0)	Y	M	.	.	.	A	C	A	.	T	.	.	C			
	14 (1)	T	A	.	.	.	.	C	A	T	K	.	.	C			
	15 (1)	T	M	.	.	.	A	Y	.	Y	.	.	.	T			
	16 (1)	T	.	.	.	.	R	Y	.	Y	.	.	.	.			
	17 (1)	Y	.	.	.	.	.	.	.	.	.	.	.	T			
	18 (1)	T	A	.	.	.	A	C	A	T	.	.	.	C			
	19 (1)	T	M	.	.	.	A	.	.	.	.	.	.	C			
	20 (1)	T	M	.	.	.	.	Y	.	Y	.	.	.	.			
	21 (1)	Y	M	G	.	.	R	.	.	.	.	.	.	C			
	22 (1)	Y	M	.	.	.	.	Y	.	Y	.	A	.	C			
	23 (1)	Y	M	.	.	.	A	Y	R	Y	.	.	.	C			
	24 (1)	T	M	.	.	.	R	Y	R	Y	T	M	.	.			
	25 (1)	.	.	.	.	.	.	.	.	.	.	.	.	T			
	26 (1)	T	A	.	.	.	A	C	A	T	.	.	.	T			
	27 (1)	.	.	.	.	.	.	.	.	.	.	.	.	C			
<i>CaVPS13</i>		33	49	134	136	212	217	241	281	282	320	322	326	328	334	370	375
	1* (0)	G	A	A	T	G	Y	A	G	G	R	T	G	R	G	A	C
	2* (0)	.	C	.	.	.	C	.	.	.	G	K	.	G	.	Y	Y
	3* (4)	.	C	.	.	.	T	.	A	.	G	G	.	G	.	T	T
	4* (0)	.	C	.	.	.	T	.	A	.	G	.	.	G	.	T	.
	5* (1)	.	C	.	.	.	T	G	.	.	G	.	.	G	.	T	.
	6* (2)	.	C	.	.	.	.	.	R	.	G	K	.	G	.	.	Y

Continued on following page

TABLE A1—Continued

Fragment	Genotype <sup>a</sup> (no. of isolates)	Nucleotide <sup>b</sup> at position:															
		33	49	134	136	212	217	241	281	282	320	322	326	328	334	370	375
<i>CaVPS13</i>																	
	7* (14)	.	C	.	.	.	.	.	R	.	G	K	.	G	.	W	Y
	8* (0)	.	C	.	.	.	.	.	R	.	G	.	.	G	.	.	.
	9* (0)	.	C	.	.	.	.	R	.	.	G	.	.	G	.	.	.
	10* (0)	.	C	.	.	.	.	R	.	.	G	K	.	G	.	.	Y
	11* (1)	.	C	W	.	.	T	.	A	.	G	.	.	G	.	T	.
	12* (0)	.	C	.	.	R	T	G	.	R	G	.	.	G	R	W	.
	13* (0)	.	C	.	.	R	.	R	.	.	G	.	.	G	.	.	.
	14* (0)	.	C	T	.	.	T	.	A	.	G	.	.	G	.	T	.
	15* (0)	.	C	W	Y	.	T	R	R	.	G	.	.	G	.	W	.
	16* (0)	.	M	.	.	.	T	.	.	.	A	.	.	G	.	.	.
	17* (0)	.	M	.	.	.	T	.	R	.	.	.	.	G	.	.	.
	18* (0)	.	M	.	.	.	T	R	.	R	.	.	.	G	.	.	.
	19* (0)	.	M	.	.	.	.	.	R	.	G	.	R	G	.	.	.
	20* (0)	R	C	.	.	R	T	R	.	.	G	.	.	.	.	.	.
	21* (2)	.	C	.	.	.	C	.	.	.	G	.	.	G	.	.	.
	22* (0)	.	C	.	.	.	.	.	R	.	G	K	.	G	.	T	Y
	23* (0)	.	M	.	.	.	T	.	.	.	.	.	.	G	.	.	.
	24* (0)	.	C	.	.	.	T	R	.	R	G	.	.	G	.	W	.
	25 (2)	.	C	.	.	.	.	.	R	.	G	G	.	G	.	W	Y
	26 (1)	.	C	.	.	.	T	R	R	.	G	K	.	G	.	T	Y
	27 (2)	.	C	.	.	.	.	.	R	.	G	K	.	G	.	W	.
	28 (1)	.	C	.	.	.	.	.	R	R	G	G	.	G	.	T	.
	29 (1)	.	M	.	.	.	T	R	.	.	.	.	.	G	.	W	.
	30 (3)	.	C	.	.	.	.	R	.	.	G	.	.	G	.	W	.
	31 (1)	.	.	.	.	.	T	R	.	.	A	.	.	G	.	.	.
	32 (1)	.	C	.	.	.	T	.	A	.	G	K	.	G	.	T	T
	33 (3)	.	.	.	.	.	T	.	.	.	A	.	.	G	.	.	.
	34 (1)	.	C	.	.	.	.	.	.	A	.	.	.	G	.	.	.
	35 (1)	.	C	.	.	.	.	.	R	.	G	K	.	G	.	W	T
	36 (1)	.	C	.	.	.	T	.	R	.	G	K	.	G	.	T	Y
	37 (1)	.	C	.	.	.	T	.	.	.	G	K	.	G	.	W	Y
	38 (1)	.	M	.	.	.	T	R	.	R	.	.	.	G	.	W	.
	39 (1)	.	C	.	.	.	T	G	.	.	.	.	.	G	.	T	.
	40 (1)	.	M	.	.	.	T	.	R	.	.	.	.	G	.	W	.
	41 (1)	.	C	.	.	.	T	.	R	.	G	.	.	G	.	W	.
	42 (1)	.	C	.	.	.	T	R	R	.	G	K	.	G	.	T	.
	43 (1)	A	C	.	.	.	T	.	.	.	G	K	.	G	.	W	Y
	44 (1)	.	C	T	.	.	T	.	A	.	G	G	.	G	.	T	.
<i>ZWF1</i>		23	31	43	49	55	337	379	482								
	1 (3)	A	T	T	T	T	T	T	A								
	2 (18)	.	Y	W	.	.	Y	Y	W								
	3 (4)	.	C	A	.	.	C	C	T								
	4 (1)	.	Y	W	.	.	Y	C	W								
	5 (3)	M	Y	.	Y	Y	Y	.	.								
	6 (2)	C	C	.	C	C	C	.	.								
	7 (1)	.	C	W	.	Y	.	.	.								
	8 (1)	.	C	W	.	.	Y	Y	W								
	9 (1)	M	.	W	Y	Y	Y	.	.								
	10 (1)	M	C	W	C	C	C	.	.								
	11 (1)	.	Y	W	.	.	Y	Y	T								
	12 (1)	M	C	.	Y	Y	C	Y	W								
	13 (1)	M	C	W	Y	Y	C	Y	W								
	14 (1)	.	.	.	.	.	C	C	T								
	15 (2)	.	Y	A	.	.	Y	Y	W								
	16 (2)	.	Y	W	Y	Y	Y	.	.								
	17 (1)	M	Y	.	Y	.	Y	.	.								
	18 (1)	.	Y	W	.	.	.	Y	W								
	19 (1)	.	C	A	.	.	C	Y	W								
	20 (2)	.	Y	A	.	.	Y	Y	T								
	21 (1)	.	C	A	Y	Y	C	Y	W								
	22 (1)	C	.	.	C	C	C	.	.								

<sup>a</sup> \*, genotype according to Bounoux et al. (2).  
<sup>b</sup> R, A or G; Y, C or T, K, G or T; M, A or C; S, G or C; and W, A or T.

## ACKNOWLEDGMENTS

For the pilot phase of this study, A. Tavanti was supported by the University of Pisa, and the study was generously supported by an unrestricted grant from Pfizer UK. Our MLST research is now supported by grant 069615 from the Wellcome Trust. We also acknowledge the British Society for Antimicrobial Chemotherapy and the BBSRC for laboratory support.

We thank Amanda Davidson for excellent technical assistance; Merck, Inc., for strain T26; and the colleagues who have supplied clinical isolates of *C. albicans* to our collection over the last 30 years.

## REFERENCES

1. Arnaviel, S., T. De Meus, A. Blancard, M. Mallie, F. Renaud, and J. M. Bastide. 2000. Multicentric genetic study of *Candida albicans* isolates from non-neutropenic patients using multilocus enzyme electrophoresis typing: population structure and mode of reproduction. *Mycoses* **43**:109–117.
2. Bougnoux, M.-E., S. Morand, and C. d'Enfert. 2002. Usefulness of multilocus sequence typing for characterization of clinical isolates of *Candida albicans*. *J. Clin. Microbiol.* **40**:1290–1297.
3. Brem, R. B., G. Yvert, R. Clinton, and L. Kruglyak. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**:752–755.
4. Chen, Y. C., J. D. Eisner, M. M. Kattar, S. L. Rassoulia-Barrett, K. LaFe, S. L. Yarfitz, A. P. Limaye, and B. T. Cookson. 2000. Identification of medically important yeasts using PCR-based detection of DNA sequence polymorphisms in the internal transcribed spacer 2 region of the rRNA genes. *J. Clin. Microbiol.* **38**:2302–2310.
5. Dingle, K. E., F. M. Colles, D. R. A. Wareing, R. Ure, A. J. Fox, F. E. Bolton, H. J. Bootsma, R. J. L. Willems, R. Urwin, and M. C. Maiden. 2001. Multilocus sequence typing system for *Campylobacter jejuni*. *J. Clin. Microbiol.* **39**:14–23.
6. Douglas, C. M., J. A. Dippolito, G. J. Shei, M. Meinz, J. Onishi, J. A. Marrinan, W. Li, G. K. Abruzzo, A. Flattery, K. Bartizal, A. Mitchell, and M. B. Kurtz. 1997. Identification of the *FKSI* gene of *Candida albicans* as the essential target of 1, 3- $\beta$ -D-glucan synthase inhibitors. *Antimicrob. Agents Chemother.* **41**:2471–2479.
7. Enright, M. C., N. P. J. Day, C. E. Davies, S. J. Peacock, and B. G. Spratt. 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J. Clin. Microbiol.* **38**:1008–1015.
8. Enright, M. C., B. G. Spratt, A. Kalia, J. H. Cross, and D. E. Bessen. 2001. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between emm type and clone. *Infect. Immun.* **69**:2416–2427.
9. Feil, E. J., and B. G. Spratt. 2001. Recombination and the population structures of bacterial pathogens. *Annu. Rev. Microbiol.* **55**:561–590.
10. Forche, A., G. Schonian, Y. Graser, R. Vilgalys, and T. G. Mitchell. 1999. Genetic structure of typical and atypical populations of *Candida albicans* from Africa. *Fungal Genet. Biol.* **28**:107–125.
11. Graser, Y., M. Volovsek, J. Arrington, G. Schonian, W. Presber, T. G. Mitchell, and R. Vilgalys. 1996. Molecular markers reveal that population structure of the human pathogen *Candida albicans* exhibits both clonality and recombination. *Proc. Natl. Acad. Sci. USA* **93**:12473–12477.
12. Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**:557–580.
13. Hull, C. M., and A. D. Johnson. 1999. Identification of a mating type-like locus in the asexual pathogenic yeast *Candida albicans*. *Science* **285**:1271–1275.
14. Hunter, P. R. 1991. A critical review of typing methods for *Candida albicans* and their applications. *Crit. Rev. Microbiol.* **17**:417–434.
15. Kurtz, M. B., G. Abruzzo, K. Bartizal, J. A. Marrinan, W. Li, J. Milligan, K. Nollstadt, and C. M. Douglas. 1996. Characterization of echinocandin-resistant mutants of *Candida albicans* — genetic, biochemical, and virulence studies. *Infect. Immun.* **64**:3244–3251.
16. Lockhart, S. R., J. J. Fritch, A. S. Meier, K. Schroppel, T. Srikantha, R. Galask, and D. R. Soll. 1995. Colonizing populations of *Candida albicans* are clonal in origin but undergo microevolution through C1 fragment reorganization as demonstrated by DNA fingerprinting and C1 sequencing. *J. Clin. Microbiol.* **33**:1501–1509.
17. Lockhart, S. R., C. Pujol, K. J. Daniels, M. G. Miller, A. D. Johnson, M. A. Pfaller, and D. R. Soll. 2002. In *Candida albicans*, white-opaque switchers are homozygous for mating type. *Genetics* **162**:737–745.
18. Lott, T. J., and M. M. Effat. 2001. Evidence for a more recently evolved clade within a *Candida albicans* North American population. *Microbiology* **147**:1687–1692.
19. Maiden, M. C. J., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* **95**:3140–3145.
20. Maynard Smith, J., N. H. Smith, M. O'Rourke, and B. G. Spratt. 1993. How clonal are bacteria? *Proc. Natl. Acad. Sci. USA* **90**:4384–4388.
21. McCullough, M., K. V. Clemons, and D. A. Stevens. 1999. Molecular epidemiology of the global and temporal diversity of *Candida albicans*. *Clin. Infect. Dis.* **29**:1220–1225.
22. McCullough, M. J., K. V. Clemons, and D. A. Stevens. 1999. Molecular and phenotypic characterization of genotypic *Candida albicans* subgroups and comparison with *Candida dubliniensis* and *Candida stellatoidea*. *J. Clin. Microbiol.* **37**:417–421.
23. Nallapareddy, S. R., R. W. Duh, K. V. Singh, and B. E. Murray. 2002. Molecular typing of selected *Enterococcus faecalis* isolates: Pilot study using multilocus sequence typing and pulsed-field gel electrophoresis. *J. Clin. Microbiol.* **40**:868–876.
24. Odds, F. C., C. C. Kibbler, E. Walker, A. Bhamra, H. G. Prentice, and P. Noone. 1989. Carriage of *Candida* species and *C. albicans* biotypes in patients undergoing chemotherapy or bone marrow transplantation for hematological disease. *J. Clin. Pathol.* **42**:1259–1266.
25. Pujol, C., S. Joly, B. Nolan, T. Srikantha, and D. R. Soll. 1999. Microevolutionary changes in *Candida albicans* identified by the complex Ca3 fingerprinting probe involve insertions and deletions of the full-length repetitive sequence RPS at specific genomic sites. *Microbiology UK* **145**:2635–2646.
26. Pujol, C., J. Reynes, F. Renaud, M. Raymond, M. Tibayrenc, F. Ayala, F. Janbon, M. Mallie, and J. Bastide. 1993. The yeast *Candida albicans* has a clonal mode of reproduction in a population of infected human immunodeficiency virus-positive patients. *Proc. Natl. Acad. Sci. USA* **90**:9456–9459.
27. Rustad, T. R., D. A. Stevens, M. A. Pfaller, and T. C. White. 2002. Homozygosity at the *Candida albicans* MTL locus associated with azole resistance. *Microbiology-SGM* **148**:1061–1072.
28. Sadhu, C., M. J. McEachern, E. P. Rustchenko-Bulgac, J. Schmid, D. R. Soll, and J. B. Hicks. 1991. Telomeric and dispersed repeat sequences in *Candida* yeasts and their use in strain identification. *J. Bacteriol.* **173**:842–850.
29. Scherer, S., and D. A. Stevens. 1988. A *Candida albicans* dispersed, repeated gene family and its epidemiological applications. *Proc. Natl. Acad. Sci. USA* **85**:1452–1456.
30. Schonian, G., A. Forche, H. J. Tietz, M. Muller, Y. Graser, R. Vilgalys, T. G. Mitchell, and W. Presber. 2000. Genetic structure of geographically different populations of *Candida albicans*. *Mycoses* **43**:51–56.
31. Schroppel, K., M. Rotman, R. Galask, K. MAC, and D. R. Soll. 1994. Evolution and replacement of *Candida albicans* strains during recurrent vaginitis demonstrated by DNA fingerprinting. *J. Clin. Microbiol.* **32**:2646–2654.
32. Soll, D. R. 2000. The ins and outs of DNA fingerprinting the infectious fungi. *Clin. Microbiol. Rev.* **13**:332–370.
33. Tamura, M., K. Watanabe, Y. Mikami, K. Yazawa, and K. Nishimura. 2001. Molecular characterization of new clinical isolates of *Candida albicans* and *C. dubliniensis* in Japan: analysis reveals a new genotype of *C. albicans* with group I intron. *J. Clin. Microbiol.* **39**:4309–4315.
34. Tibayrenc, M. 1997. Are *Candida albicans* natural populations subdivided? *Trends Microbiol.* **5**:253–257.
35. White, T. C. 1997. Increased mRNA levels of *ERG16*, *CDR*, and *MDR1* correlate with increases in azole resistance in *Candida albicans* isolates from a patient infected with human immunodeficiency virus. *Antimicrob. Agents Chemother.* **41**:1482–1487.