Molecular Genotyping of *Candida parapsilosis* Group I Clinical Isolates by Analysis of Polymorphic Microsatellite Markers

Brent A. Lasker, Geraldine Butler, and Timothy J. Lott

*Mycotic Diseases Branch, Division of Bacterial and Mycotic Diseases, National Centers for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, 30333, and Department of Biochemistry, Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin 4, Ireland*

Received 1 April 2005/Returned for modification 7 June 2005/Accepted 13 December 2005

*Candida parapsilosis*, a pathogenic yeast, is composed of three newly designated genomic species that are physiologically and morphologically indistinguishable. Nosocomial infections caused by group I *C. parapsilosis* are often associated with the breakdown of infection control practices and the contamination of medical devices, solutions, and indwelling catheters. Due to the low levels of nucleotide sequence variation that are observed, an investigation of the size polymorphisms in loci harboring microsatellite repeat sequences was applied for the typing of *C. parapsilosis* group I isolates. Following amplification by PCR, the size of each amplification product was determined automatically by capillary electrophoresis. A total of 42 *C. parapsilosis* group I isolates were typed by microsatellite analysis, and their profiles were compared to the hybridization profiles obtained by use of the Cp3-13 DNA probe. A high degree of discrimination (discriminatory power = 0.971) was observed by microsatellite analysis. The number of different alleles per locus ranged from 14 for locus B to 5 for locus C. Microsatellite analysis detected 30 different microsatellite genotypes, with 24 genotypes represented by a single isolate. Comparison of the genotypes obtained by microsatellite analysis and those obtained by analysis of the Cp3-13 hybridization profiles showed that they were similar, and these methods were able to identify related and unrelated isolates. Some discrepancies were observed between the methods and may be due to higher mutation rates and/or homoplasy by microsatellite markers. Identical results were observed between microsatellite analysis and Cp3-13 DNA hybridization profile analysis for *C. parapsilosis* isolates obtained from two patients, demonstrating the reproducibilities of the methods in vivo. Identical microsatellite profiles were observed for isolates displaying different phenotypic switching morphologies. Indistinguishable Cp3-13 DNA hybridization profiles were observed for six epidemiologically related isolates; however, only three of six primary isolates had identical microsatellite profiles. Size variation at a single locus was observed for three of six isolates obtained either after the outbreak period or from a different body site, suggesting the potential of the method to detect microevolutionary events. Interestingly, for most loci a single allele per strain was observed; in contrast, two alleles per locus were observed for some strains, and consistent with the findings for natural isolates, some isolates may be aneuploid. Due to the potential for high throughput, reproducibility, and discrimination, microsatellite analysis may provide a robust and efficient method for the genotyping of large numbers of *C. parapsilosis* group I isolates.

*Candida* species have been reported to be responsible for approximately 10% of all nosocomial bloodstream infections occurring in the United States and the fourth most common pathogen causing nosocomial bloodstream infections (8, 20, 42). Among the *Candida* species causing nosocomial infections, the opportunistic yeast pathogen *Candida parapsilosis* is frequently isolated. For instance, in some institutions in Latin America, Canada, and Asia, *C. parapsilosis* is currently considered the second or third most common species of yeast isolated from blood cultures (40, 49). This yeast has been reported to be responsible for a broad variety of clinical manifestations, including fungemia, endocarditis, endophthalmitis, peritonitis, and infectious arthritis (55). *C. parapsilosis* infections generally occur in individuals with impaired immune systems, neutropenia, or burns and in individuals in neonatal or surgical intensive care units (33, 35, 40, 43, 55).

*C. parapsilosis* has been isolated from several environmental sources, including soil and seawater, and from epithelial and mucosal surfaces, skin, and nails, where it is normally considered part of the benign commensal flora of humans and mammals (6, 10, 55). In contrast to *Candida albicans*, infections by *C. parapsilosis* may occur without prior colonization of the patients, especially in infant populations (28, 51). A common denominator for several outbreaks of *C. parapsilosis* infections is the breakdown of infection control practices by health care workers, which leads to the contamination of intravascular catheters and other medical devices (22, 28, 33). *C. parapsilosis* has been isolated from the hands of health care workers who install and maintain these medical devices, suggesting a potential route for transmission (28, 52). Other physiological factors believed to be important for colonization or transmission include secretory aspartyl-proteinase production (6, 24), as well as adhesion to medical materials, slime production, and the ability to form biofilms (22, 24, 41).

Isolates of *C. parapsilosis* have been reported to be physiologically indistinguishable but genetically heterogeneous. Investigations have suggested that *C. parapsilosis* is a complex composed of three genetically distinct groups, based on ran-
domly amplified polymorphic DNA (RAPD) analysis, isoenzyme analysis, nucleotide sequence analysis (21, 27, 30, 36), and DNA-DNA hybridization (46). Recently, representative isolates of the three groups were analyzed by multilocus sequence typing (MLST) by two independent groups (13, 53). Tavanti et al. (53) proposed that each of the three groups of C. parapsilosis be considered a new species, based on the high degree of sequence variation observed between groups. Group I isolates were proposed to retain the name C. parapsilosis, whereas group II isolates were given the species name of Candida orthopsilosis, and likewise, group III isolates were given the species name of Candida metapsilosis (53). Of the three groups, most of the clinical isolates are group I isolates, which may be partially due to their enhanced ability to form biofilms (24). The low degree of sequence variation observed for group I isolates suggests that they emerged more recently than group II and III isolates (13, 53). Fundyga et al. (13) also reported a wide variation of genome size between isolates, suggesting that natural isolates may be predominantly aneuploid (>1n but <2n).

Several methods have previously been used to type and distinguish isolates of C. parapsilosis at the molecular level in order to determine routes of transmission, strain persistence, and sources during outbreaks or relapses. Both isoenzyme analysis (27) and digestion of genomic DNA embedded in agarose plugs with restriction endonucleases with low frequencies of digestion were limited by low degrees of discrimination (41). Electrophoretic karyotype analysis has frequently been used and has been demonstrated to have a high degree of discrimination, but the method lacks the ability to accurately determine the degree of genetic relatedness between strains (6, 32, 33, 45, 51). RAPD analysis has been widely used for strain typing (32, 33, 43). While RAPDs are generally available to many researchers, to date, there are no standardized sets of primers, isolates, or amplification conditions; and a more serious problem is the relative instability of RAPD profiles (1). The results obtained with a complex DNA probe, Cp3-13, which has been reported to have a high degree of discrimination and to be able to group isolates, were found to be in good agreement with those of RAPD analysis by the use of six different RAPD primers (9). By MLST, a low degree of nucleotide sequence diversity was observed by two independent investigations with group I isolates (13, 53). Analysis by MLST may be limited, since low levels of nucleotide variation may pose problems for genetic analysis (54), therefore necessitating the search for more polymorphic markers that can be used to discriminate between isolates. We therefore investigated the use of a method that uses polymorphic microsatellite markers (PMMs) for the typing and analysis of C. parapsilosis group I isolates. Microsatellites are defined as short 2- to 10-bp multiple tandem repeats and are increasing in utility and importance as genetic markers. Analysis of microsatellite loci has an advantage over other commonly used typing methods since microsatellite loci behave as codominant markers, evolve rapidly in a genome, and may be able to distinguish between isolates for microorganisms with low degrees of sequence variation.

The goals of this investigation were to identify and evaluate polymorphic microsatellite loci obtained from group I C. parapsilosis genomic sequences for use as genetic markers to discriminate between isolates. The performance of microsatellite analysis was determined by comparison of the results to the typing results obtained by Cp3-13 DNA hybridization profile analysis. Microsatellite analysis demonstrated high degrees of both discriminatory power and reproducibility, and the ability to detect microevolutionary variations of isolates obtained from different body sites suggests its potential utility as an important adjunct for outbreak and epidemiologic investigations.

### MATERIALS AND METHODS

**C. parapsilosis isolates.** The 42 isolates of C. parapsilosis group I used in the present investigation and their sources are listed in Table 1. P. Lehmann (Medical College of Ohio) kindly provided C. orthopsilosis isolate MCO471 (ATCC 96140; group II C. parapsilosis); C. metapsilosis MCO429 (ATCC 96143; group III C. parapsilosis); and C. parapsilosis isolates MCO478 (ATCC 22019), MCO439, and MCO441. These isolates have also been described by Lin et al.
TABLE 2. Characteristics of seven C. parapsilosis PMMs

<table>
<thead>
<tr>
<th>Locus</th>
<th>Repeat type</th>
<th>Primer sequences (5’ to 3’)</th>
<th>Range of PCR product size (bp)</th>
<th>No. of alleles</th>
<th>Genotypic frequenciesa</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(CA)25</td>
<td>AGGCGCTCAATCTAAGACAGAGAGCT</td>
<td>103–111</td>
<td>6</td>
<td>0.024–0.309</td>
</tr>
<tr>
<td>B</td>
<td>(CA)30</td>
<td>AGGTTTGATGTTGCTCATTGGTCTT</td>
<td>100–149</td>
<td>14</td>
<td>0.024–0.262</td>
</tr>
<tr>
<td>C</td>
<td>(GT)22</td>
<td>AGCTGGACCTTGATCAAGTCTGCT</td>
<td>126–150</td>
<td>4</td>
<td>0.071–0.595</td>
</tr>
<tr>
<td>D</td>
<td>(GT)21</td>
<td>GTGATTGCTAGTGGCTGATGTTT</td>
<td>122–143</td>
<td>6</td>
<td>0.024–0.690</td>
</tr>
<tr>
<td>E</td>
<td>(GT)18</td>
<td>ACTCATGGCGATGGCTGCAAAGG</td>
<td>148–160</td>
<td>5</td>
<td>0.024–0.452</td>
</tr>
<tr>
<td>F</td>
<td>(GT)17</td>
<td>GTGAAAGAAGGTGCTGACAGA</td>
<td>129–149</td>
<td>6</td>
<td>0.024–0.381</td>
</tr>
<tr>
<td>G</td>
<td>(CAA)14</td>
<td>TGGACCCGTTGCTGAAATGTTG</td>
<td>107–146</td>
<td>5</td>
<td>0.024–0.833</td>
</tr>
</tbody>
</table>

a Repeat based on DNA sequence for C. parapsilosis ATCC 22019.

b Frequencies of the least and the most common genotypes, respectively, for a given locus.
For these seven polymorphic loci, a single PCR amplification product was observed for the majority of isolates. In several cases, even for samples obtained from different DNA preparations from the same isolate, two ethidium bromide-stained bands were observed per isolate. For example, a conserved PCR fragment at 128 bp was observed for all eight isolates, whereas only three of the isolates displayed a second PCR fragment at 150 bp, as shown in Fig. 1C. The seven PMM and repeat types are shown in Table 2. Of the seven sequences, two contained (CA)n repeats (loki A and B), four loci contained (GT)n repeats (loki C, D, E, and F), and locus G consisted of a (CAA)14 repeat. Both the BLAST and the BLASTX programs were used to search the GenBank database for nucleotide and amino acid sequence identities, respectively. Significant sequence identity was not detected for locus A, B, C, D, or F. However, a significant degree of amino acid identity (e=140) was detected for a conserved domain located near locus E for a meiotic recombination protein, the DLH1 locus of C. albicans (GenBank accession no. U39808). Sequences adjacent to locus G detected high amino acid identity (e=173) to a domain of an open reading frame in C. albicans strain SC5314 (GenBank accession no. EAK94743) (data not shown).

The PCR fragment size was automatically determined by capillary electrophoresis with GeneScan software with the incorporation of an internal size standard into each sample. Capillary electrophoresis detected a single band per locus for the majority of isolates. Many of the bands were frequently composed of less intense stutter bands. These bands are PCR products a few base pairs shorter than the main peak and are due to slippage errors of 1 or 2 bp by Taq DNA polymerase during replication (15). The largest peak was used for precise determination of the allele size, and a representative profile for locus B for C. parapsilosis MCO478 is shown in Fig. 2A. Whereas a single PCR product was observed for the majority of isolates, of interest was the observation of two different-sized fragments (n = 25 of 294 fragments) for several loci in several isolates. Isolates displaying two different-sized fragments per locus ranged from one isolate each for loci A and E to five isolates for locus B. One example showing two different-sized fragments for C. parapsilosis MCO439 is shown in Fig. 2B. The observation of two alleles per isolate may be explained by cross-contamination of DNA preparations, but this explanation is unlikely since identical results were also observed after different DNA preparations from an isolate were used. The presence of two bands may possibly be due to the presence of pseudogenes, gene duplication, and/or perhaps a diploid genome. Alternatively, a more likely explanation is consistent with the presence of an aneuploid genome in the isolates (13). PCR products were generated for all 42 C. parapsilosis group I isolates at seven loci, resulting in 100% typeability. In comparison, not all loci were amplified for C. orthopsilosis or C. metapsilosis (data not shown), indicating that the primers are not species specific. For C. orthopsilosis MCO471, PCR products were observed from loci A, C, D, E, F, and G but not from locus B. No PCR products were observed for DNA obtained from C. metapsilosis strain MCO429 for loci A, B, and D; but PCR products were observed for loci C, E, F, and G.

In general, loci with the greatest number of repeats showed the highest degree of discriminatory power and/or number of

---

**FIG. 1.** Ethidium bromide-stained PCR fragments separated by MetaPhor agarose gel electrophoresis for eight isolates of C. parapsilosis. Lane 1, B-6244; lane 2, B-6213; lane 3, B-6212; lane 4, B-6185; lane 5, B-6183; lane 6, B-6191; lane 7, B-6190; and lane 8, MCO478. (A) Amplification of a 96-bp PCR fragment with primer set I; (B) detection of size length polymorphisms following amplification with primer set B; (C) detection of two different-sized alleles of locus C from three isolates following PCR amplification.

**RESULTS**

The FindPatterns program was used to screen 5,374 contigs, consisting of >15,000 individual reads for C. parapsilosis genomic DNA (31). Screening detected a total of 95 putative sequences containing microsatellite repeats with from 9 to 30 dinucleotide repeat units and 7 to 14 trinucleotide repeat units. Sequences were selected for preliminary analysis based only on the presence of more than eight repeat units and the availability of flanking DNA adjacent to the repeat sequences. Fifty sequences met these two criteria. PCR primer pairs were designed and synthesized for each sequence. PMMs were initially identified by electrophoresis of the PCR amplification products through MetaPhor agarose gels. No size length polymorphisms were observed for the majority of loci (43 of 50) by MetaPhor agarose gel electrophoresis. An example is shown in Fig. 1A. The lack of fragment size polymorphisms in the PCR amplification products for 6 of 43 different loci was confirmed by capillary electrophoresis (data not shown). However, we did observe fragment size polymorphisms for PCR products resolved by MetaPhor agarose gel electrophoresis for seven loci. The fragment size differences are shown in Fig. 1B for locus B.
alleles (Table 2). The highest degrees of discriminatory power (D) were observed for locus A (D = 0.837) and locus B (D = 0.876). Analyses with loci A and B detected 6 and 14 different alleles, respectively. The most common allele for locus A was at 109 bp (n = 12), and the most common allele for locus B was at 147 bp (n = 11). Of the seven loci, locus G showed the lowest degree of discriminatory power (D = 0.341). By combining the results for all seven loci together, a relatively high degree of discrimination (D = 0.971) was observed by microsatellite analysis. Twenty-five microsatellite genotypes were represented by single isolates. The genetic distance among C. parapsilosis isolates is shown in Fig. 3 as an unrooted tree.

The performance and capacity to genotype isolates by microsatellite analysis were compared to the performance and capacity to genotype isolates by Cp3-13 DNA hybridization profile analysis for 41 C. parapsilosis group I isolates. As shown in Table 1, microsatellite analysis and analysis by the use of Cp3-13 DNA hybridization profiles were both found to have a high capacity to discriminate among isolates, and a relatively high level of agreement of the results of the two methods was displayed. For instance, microsatellite analysis detected 30 different genotypes, whereas Cp3-13 DNA hybridization analysis detected 31 different genotypes; and all 24 microsatellite genotypes represented by single unrelated isolates were confirmed by analysis with Cp3-13 DNA. By using a simple matching coefficient, isolates typed by Cp3-13 DNA hybridization profile analysis and microsatellite analysis were 80% and 87% concordant, respectively, when microevolutionary variations in epidemiologically related isolates are considered. The genotypes for eight isolates were not concordant (Table 1). Three isolates from cultures of hand wash specimens obtained from individuals involved in the community outbreak in Mississippi (isolates B-6191, B-6191b, and B-6192) were indistinguishable by microsatellite analysis, but each isolate was considered a different strain by Cp3-13 DNA hybridization analysis. Isolates B-6187, B-6225, and B-6245 were identical by microsatellite analysis but shared the Cp3-13 hybridization profiles with isolates B-6183, B-6187, and B-6190. The microsatellite genotype observed for the isolate from the blood of a patient from Louisiana (isolate Y-546-91) was shown to be different from the microsatellite genotype observed for two isolates, isolates Y-532-91 and Y-542-91, obtained from a second patient. In comparison, these three isolates were indistinguishable by Cp3-13 DNA hybridization profile analysis. Isolate B-6213, an isolate from California, and isolate B-6204, obtained from hand washes from individuals in Mississippi, were identical by microsatellite analysis but were unrelated by Cp3-13 DNA hybridization analysis.

Comparison of the clustering and the distribution of the isolates obtained by the two methods of typing showed good agreement (Fig. 3 and 4). For example, the six Cp3-13 type 10 isolates (Table 1) obtained from an outbreak in Mississippi were assigned to two closely related microsatellite types, types 11 and 13. Likewise, epidemiologically related isolates from Georgia and Louisiana were similarly clustered by both methods. However, some discrepancies were observed. Cp3-13 hybridization profile analysis demonstrated a more robust ability to distinguish among isolates than microsatellite analysis, which may be due to the combination of conserved and rapidly evolving markers in a complex DNA probe such as Cp3-13 and its higher level of discriminatory power (9). Microsatellites appear to be evolving with a significantly higher rate of sequence divergence than Cp3-13 DNA. This faster rate of sequence divergence by microsatellites, as reflected in differences in size polymorphisms, may help to drive the more rapid establishment of unrelated profiles, which can be useful in outbreak situations but less effective for the determination of long-term genetic relatedness.

The reproducibility and stability of the microsatellite profiles were determined by two methods. The first method was repeated analysis of the microsatellite genotypes by using the same or different DNA preparations obtained from the same isolate. In both cases, identical microsatellite genotypes were observed for independent DNA preparations and repeated analysis of the same DNA preparation. In vivo reproducibility was examined by comparing the microsatellite genotypes for isolates obtained from the same patient or from epidemiologically related isolates. The types of three primary isolates obtained from a single patient (isolates Y-346-90, Y-350-90, and Y-351-90) were identical by both microsatellite analysis and Cp3-13 DNA hybridization analysis (Table 1). Likewise, analysis of two isolates showing phenotypic switching (32) (isolates Y-532-91 and Y-542-91) and two isolates obtained from a patient in California (isolates B-6215 and B-6216) were shown to be identical by both Cp3-13 DNA hybridization analysis and microsatellite analysis. Six epidemiologically linked clinical isolates (isolates B-6178, B-6183, B-6186, B-6187, B-6225, and B-6245) and three isolates from hand wash specimens (isolates B-6191, B-6191b, and B-6192) were obtained from individuals during or following an outbreak of bloodstream infections at a community hospital in Mississippi (3, 24). The six clinical isolates were all indistinguishable by analysis of the Cp3-13 DNA hybridization profiles, but their profiles differed from the Cp3-13 DNA profiles obtained for isolates from the hands of
health care workers. Identical microsatellite profiles were observed for three patient isolates (isolates B-6178, B-6183, and B-6186) and the three hand wash isolates obtained during the outbreak (Table 1). Isolates B-6187, B-6225 and B-6245, obtained from either a different body site or after the outbreak, had a microsatellite profile different from that observed for the case isolates, as shown in Fig. 3. The difference was a single change in the fragment size in locus A, with CA dinucleotide repeats, from 109 bp to 107 bp. Likewise, six of the seven microsatellite loci were identical between isolate B-6225, which was obtained from sputum, and the case isolates. Size variation at a single locus is consistent for a strain that has undergone a microevolutionary change.

**DISCUSSION**

To our knowledge this is the first report identifying and characterizing PMMs as a new class of genetic markers for use in the molecular subtyping of *C. parapsilosis* group I isolates. Microsatellites are found in all genomes and are assuming greater importance as molecular markers. They have been widely used for molecular typing and genetic analysis of fungal populations. Several highly variable PMMs were successfully used to characterize and rapidly type isolates of *C. albicans* with a high degree of discriminatory power and reproducibility (2, 47). Microsatellite analysis has been used for the genotyping of *C. albicans* isolates from healthy individuals (5), human immunodeficiency virus-infected patients (34), and patients with recurrent vulvovaginal infections (47). Ohst et al. (37) used microsatellites to examine the strain distribution with respect to geographic origin for 130 strains of the dermatophytes *Trichophyton rubrum* and *Trichophyton violaceum*. Microsatellites have been useful for the typing of a variety of fungi, such as *Aspergillus fumigatus* (1), *Saccharomyces cerevisiae* (16), and *Coccidioides immitis* (12). In *Penicillium marneffei*, a dimorphic pathogenic fungus with a low degree of nucleotide sequence variation, analysis of PMMs was able to detect a high degree of genetic diversity and geographically distinct allele combinations (11, 26).

In our investigation, microsatellite analysis showed a highly reproducible and robust capacity to discriminate between epidemiologically unrelated isolates (Table 1; Fig. 3). Likewise, genotyping by the use of microsatellites was also able to identify epidemiologically related and closely related isolates, as well as show a relatively high degree of agreement with the genotypes obtained by Cp3-13 DNA hybridization analysis (80%), validating the utility of PMMs for the molecular typing of *C. parapsilosis* group I isolates. Although there was good agree-
ment between the methods for the assignment of genotypes, disagreement of clustering of unrelated isolates was often observed (Fig. 3 and 4). These conflicts may be due to several factors. Mutation rates for microsatellites are high and can range from $10^{-2}$ to $10^{-6}$ mutations per generation, whereas point mutations accumulate more slowly, at approximately $10^{-9}$ mutations per generation (4, 44). The higher rate of genomic variation than Cp3-13 DNA variation may result in the accumulation of unrelated patterns at a higher rate. Whereas microsatellites have been very useful as genetic markers for estimation of the genetic relatedness within a species, constraints on size distribution may reduce the reliability of microsatellites in studies involving phylogenetic reconstruction of more distantly related organisms (23).

Another possible concern is that the constraints on the microsatellite size distribution may lead to alleles with an identical size but not necessarily with a common ancestor, commonly known as homoplasy. Isolate B-6213, which was obtained from California, and isolate B-6204, which was obtained from Mississippi, may be examples of homoplasy, since they share a common microsatellite genotype, even though they were obtained from geographic locations separated by a long distance, suggesting that they do not share a common ancestry. To reduce the chance of detecting size variation outside the microsatellite region due to insertions and/or deletions, we designed PCR primers complementary to a target sequence as close as possible to the repeat region. Genetic relatedness may be underestimated, since nucleotide diversity may be present in regions flanking the microsatellite loci (38). In this investigation we cannot completely eliminate the pos-

FIG. 4. Cluster analysis of Cp3-13 hybridization profiles computed for 41 isolates of *C. parapsilosis* by the use of Bionumerics version 4.0 software.
sibility of homoplasy, since not all the loci were sequenced. One strategy that can be used to reduce the influence of homoplasy at a single locus is the analysis of several PMMs (12), as was done in this investigation.

Analysis of the Cp3-13 DNA hybridization profiles also has disadvantages, such as the potential for incomplete digests, the requirement for relatively large amounts of purified genomic DNA, the presence of weakly hybridizing bands, subjectivity during analysis, the inability to distinguish heterozygosity, and potential technical artifacts in alignment of profiles (50). At present, Cp3-13 DNA hybridization analysis provides a better resolution of the population structure, whereas microsatellite analysis is useful for outbreak investigations.

Microsatellite analysis offers several advantages over previously used molecular typing methods for C. parapsilosis. PMMs have a discriminatory power higher than that reported for MLST (13, 53) or multilocus enzyme analysis (30), and their discriminatory power is at least comparable to that observed by investigators using Cp3-13 DNA analysis (3, 9), RAPD analysis (9, 32, 35), and/or electrophoretic karyotyping (6, 32, 35, 51). Both RAPD analysis (1) and electrophoretic karyotyping (18) have been shown to be potentially unstable, complicating interpretation of the results. High-throughput PMM analysis should be possible, since the method requires relatively small amounts of DNA and is amenable to automation. Our investigation screened for microsatellites not from the entire genome but from the nucleotide sequences available from a genome survey of C. parapsilosis, and so we expect to identify additional PMMs from other regions of the genome.

A high degree of discrimination (D = 0.971) was obtained by combining seven PMMs. The high degree of discriminatory power is in good agreement with the discriminatory power reported in other investigations with PMMs (1, 2, 11, 26, 47) and may be due to a mutation rate for microsatellites higher than that from the accumulation of point mutations in a genome. The high mutation rates for PMMs are believed to be due to two plausible molecular mechanisms for the generation of mutations in microsatellite repeats. The most common mechanism is stepwise mutation by polymerase slippage errors during DNA replication (29) or a modified two-step mutation (7). The majority of size variations observed in microsatellite repeats were biased toward mutations resulting in the gain or the loss of one or two repeat units (49). Errors in replication are in a balance with cellular mechanisms, like mismatch repair and proofreading activities performed to correct mistakes during replication. A second mechanism that leads to microsatellite length variation is recombination events, such as unequal crossing over or gene conversion (14, 19). The majority of microsatellites are believed to occur in noncoding genomic sequences because of the potentially deleterious effects of frameshift mutations in coding regions. This is in agreement with the findings of our investigation, since database searches for loci A, B, C, and D detected no significant identity to known protein motifs. Likewise, for locus G, a trinucleotide repeat, the CAA repeats appears to be in frame with the known protein motifs. Likewise, for locus G, a trinucleotide repeat, the CAA repeats appears to be in frame with the known protein motifs.

Microsatellite analysis offers several advantages over previously used molecular typing methods for C. parapsilosis. PMMs have a discriminatory power higher than that reported for MLST (13, 53) or multilocus enzyme analysis (30), and their discriminatory power is at least comparable to that observed by investigators using Cp3-13 DNA analysis (3, 9), RAPD analysis (9, 32, 35), and/or electrophoretic karyotyping (6, 32, 35, 51). Both RAPD analysis (1) and electrophoretic karyotyping (18) have been shown to be potentially unstable, complicating interpretation of the results. High-throughput PMM analysis should be possible, since the method requires relatively small amounts of DNA and is amenable to automation. Our investigation screened for microsatellites not from the entire genome but from the nucleotide sequences available from a genome survey of C. parapsilosis, and so we expect to identify additional PMMs from other regions of the genome.

A high degree of discrimination (D = 0.971) was obtained by combining seven PMMs. The high degree of discriminatory power is in good agreement with the discriminatory power reported in other investigations with PMMs (1, 2, 11, 26, 47) and may be due to a mutation rate for microsatellites higher than that from the accumulation of point mutations in a genome. The high mutation rates for PMMs are believed to be due to two plausible molecular mechanisms for the generation of mutations in microsatellite repeats. The most common mechanism is stepwise mutation by polymerase slippage errors during DNA replication (29) or a modified two-step mutation (7). The majority of size variations observed in microsatellite repeats were biased toward mutations resulting in the gain or the loss of one or two repeat units (49). Errors in replication are in a balance with cellular mechanisms, like mismatch repair and proofreading activities performed to correct mistakes during replication. A second mechanism that leads to microsatellite length variation is recombination events, such as unequal crossing over or gene conversion (14, 19). The majority of microsatellites are believed to occur in noncoding genomic sequences because of the potentially deleterious effects of frameshift mutations in coding regions. This is in agreement with the findings of our investigation, since database searches for loci A, B, C, and D detected no significant identity to known protein motifs. Likewise, for locus G, a trinucleotide repeat, the CAA repeats appears to be in frame with the coding region and codes for a string of glutamines of the C. albicans homolog of EAK94743. Interestingly, a high degree of amino acid identity (e −140) was observed for DLH1, a meiotic recombination protein of C. albicans, and sequences adjacent to locus E. It is tempting to speculate that the loss of a meiotic pathway may have occurred. The microsatellite repeat length has been shown to be an important predictor of the rate of mutation of a given allele and is in agreement with our characterization of C. parapsilosis microsatellites. The highest degree of discriminatory power was observed for loci A and B. These two loci have the longest number of repeats: >25 repeat units for locus A and >30 repeat units for locus B. Likewise, for 50 loci found to be harboring repeats, only loci with >14 repeat units were observed to be polymorphic (Table 2), and this may be useful information for the selection of new PMMs for analysis.

The high degree of in vitro and in vivo reproducibilities may be due in part to the incorporation of internal size standards into each sample and the use of stringent conditions for PCR amplification, such as the amplification of target sequences at 58°C and the annealing of primers for a short period of time to reduce the chance of mismatch during hybridization to target sequences. In vivo reproducibility was detected by analysis of isolates obtained from a neonate in Georgia, two isolates from a patient in California, as well as two isolates whose phenotypes switched (Table 1). Likewise, microsatellite genotypes for isolates obtained either from sputum or after the outbreak period were found to differ generally by only a single dinucleotide repeat unit, as would be expected for a molecular mechanism due to DNA polymerase slippage, and suggests that the method has the ability to detect microevolution in isolates from infected patients. The differences in microsatellite profiles reflect microevolutionary changes in the genome and not marker instability, since the analysis of three hand wash isolates (isolates B-6191, B-6191b, and B-6192) by microsatellite analysis showed that they were indistinguishable; but they were observed to be unrelated by Cp3-13 DNA hybridization analysis. Microevolutionary events were first observed in C. parapsilosis isolates grown for 200 generations and for serial isolates obtained from the oral cavity of a human immunodeficiency virus-infected individual by Cp3-13 DNA hybridization analysis (9). Knowledge of the mutation rate may allow the more accurate estimation of the time of infection and genetic relatedness.

This is the first report to describe the utility of microsatellite analysis for the molecular typing of isolates of a microorganism with an aneuploid genome. In organisms with a haploid genome, a single PCR product is expected, as was shown in Fig. 2A. In contrast, in an organism with an aneuploid genome, like C. parapsilosis, with strain-specific variations in genome size, one or two PCR amplification products may be detected. Heterozygous loci were observed to be isolate specific and were useful for contributing to unique allelic combinations as well as to enhancing the discriminatory power of the method. Microsatellite profiles with two different-sized alleles have been detected, as shown in Fig. 2B, but were not observed in organisms with haploid genomes, such as A. fumigatus (1) or P. marneffei (11, 26). This provides additional support for the findings that naturally occurring isolates of C. parapsilosis may be aneuploid (13); however, gene duplication may also be responsible for a second band or the possibility of a diploid genome. PMMs are codominant and may provide more accurate genotype assignments than methods that rely on the detection of restriction fragment length polymorphisms.

We observed 100% typeability (defined as the ability to


34. Metzgar, D., A. van Belmont, D. Field, R. Haubrich, and C. Wills. 1998. Random amplification of polymorphic DNA and microsatellite genotyping of pre- and posttreatment isolates of Candida spp. from human immunodefic-


References


