A Genome Sequence Survey Shows that the Pathogenic Yeast Candida parapsilosis Has a Defective MTLα1 Allele at Its Mating Type Locus†

Mary E. Logue,1 Simon Wong,2 Kenneth H. Wolfe,2 and Geraldine Butler1*

Department of Biochemistry, Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin 4, Ireland,1 and Department of Genetics, Smurfit Institute, University of Dublin, Trinity College, Dublin 2, Ireland2

Received 25 September 2004/Accepted 22 March 2005

Candida parapsilosis is responsible for ca. 15% of Candida infections and is of particular concern in neonates and surgical intensive care patients. The related species Candida albicans has recently been shown to possess a functional mating pathway. To analyze the analogous pathway in C. parapsilosis, we carried out a genome sequence survey of the type strain. We identified ca. 3,900 genes, with an average amino acid identity of 59% with C. albicans. Of these, 23 are predicted to be predominantly involved in mating. We identified a genomic locus homologous to the MTLα mating type locus of C. albicans, but the C. parapsilosis type strain has at least two internal stop codons in the MTLα open reading frame, and two predicted introns are not spliced. These stop codons were present in MTLα1 of all eight C. parapsilosis isolates tested. Furthermore, we found that all isolates of C. parapsilosis tested appear to contain only the MTLα idiomorph at the presumptive mating locus, unlike C. albicans and C. dubliniensis. MTLα sequences are present but at a different chromosomal location. It is therefore likely that all (or at least the majority) of C. parapsilosis isolates have a mating pathway that is either defective or substantially different from that of C. albicans.

† Supplemental material for this article may be found at http://cc.asm.org/.

MATERIALS AND METHODS

Strains and media. The type strain of C. parapsilosis (CLIB214) was obtained from the Collection de Levure d’Intérêt Biotechnologique, Thiverval Grignon,
elements were removed by BLASTN searches of a database containing the CROSS_MATCH and a dedicated Perl script. Mitochondrial, rRNA, and tRNA were obtained from both ends of the insert for 8,160 clones. For submission of C. parapsilosis CLIB214 and used to generate a plasmid-based random genomic survey. High-molecular-weight DNA was prepared from yeast extract-peptone-dextrose (YPD) medium. The oligonucleotide (oligo) primer sequences and no or much weaker sequences were obtained from both ends of the insert for 8,160 clones. For submission of high-quality reads to GenBank, base calling was carried out by using PHRED (error rate cutoff of 0.05 and a PHRED score of 15) and vector clipping using CROSS_MATCH and a dedicated Perl script. Mitochondrial, rRNA, and tRNA elements were removed by BLASTN searches of a database containing the equivalent sequences from Saccharomyces cerevisiae and C. albicans and the mitochondrial genome from C. parapsilosis (33). Any read with a significant expect value (E-value) of <1e-3 was excluded from subsequent analysis. Contigs were assembled by using PHRAP. Approximately 100 high-quality reads with a significant match (E < 1e-10) to Pseudomonas sequences and no or much weaker match to fungal sequences were also removed, since these were assumed to be derived from a low-level contamination of the genomic library. PCR analysis of two of these sequences confirmed that they were not present in the C. parapsilosis genome. The sequence reads have been deposited in GenBank with accession numbers CZ278593 to CZ293207.

Putative identities were assigned to many of the C. parapsilosis sequences by using a comparison to a recent annotation of the S. cerevisiae genome containing 5,616 ORFs (18). The S. cerevisiae protein with the strongest BLASTX score was considered to be the ortholog, providing that the E-value for this hit was <1e-10 and more than 1e10 lower than the E-value for the second-best hit. For each ortholog the cellular function in S. cerevisiae was determined by using the "cellular role" categories of the Yeast Proteome Database of the Incyte Bioknowledge Library (12).

Comparisons with C. albicans proteins were determined by using a haploid gene set of 7,677 ORFs (including small and overlapping ORFs) predicted from assembly 19 (16). Classification of C. albicans proteins into functional categories was performed by using the same criteria as for C. parapsilosis. Sequences that did not match the C. albicans genome with BLASTX scores of <1e-3 were compared to the nonredundant database.

For the calculation of average sequence identity, only orthologs that are reciprocal best hits were used. This corresponded to 3,547 genes compared between C. parapsilosis and C. albicans and 5,511 genes compared between C. dublinensis and C. albicans. The C. dublinensis gene sequences were obtained from the Sanger Institute (assembly of 1 May 2004 [http://www.sanger.ac.uk/Projects/C_dublinensis/]).

The C. tropicalis MTL sequences were obtained from assembly 1 at the MIT sequencing project (http://www.broad.mit.edu). We did not analyze the ORFs in detail since there appear to be several frameshifts and possible sequencing errors in the current assembly.

Isolation of MTL locus. The sequence of the entire insert from clones cp5f11 (from CaKz to Cp19.3102) and cp12f03 (from Cp19.3192 to CpGp1; see Fig. 1) was determined by AGOWA. Oligos CP4 (designed from the CpPKa ortholog) and CP2 (from the C. albicans ortholog) were used with Expand Long-Template DNA polymerase (Roche) to amplify a region of 6.3 kb between the two inserts. The PCR product was purified by using a QiAquick kit (Qiagen) according to the manufacturer's instructions, and the sequence was determined directly by GATC (Germany). The sequence of the PCR product was combined with the overlapping sequence of the cp5f11 and cp12f03 inserts to give a 15.7-kb sequence that includes the entire predicted CpMTL1 idiomorph. The sequence has been deposited with GenBank (accession number AY791981). For restriction analysis of the MTL locus, fragments of 14 to 21 kb were amplified by PCR by using Expand Long-Template DNA polymerase (Roche) and oligos 3190_1 (designed from within the Cp19.3109 ORF), CP6 (designed from sequence 900 bp upstream from the start of CgGp41), CP3 (designed from within the CpGp7), and CcCp1_1 (designed from within the CpCc77 ortholog). High-molecular-weight genomic DNA was prepared by using a Genomic-Tip 20/G kit (Qiagen).

To confirm the mutations in MTL1a, a 200-bp fragment of the ORF was amplified from eight C. parapsilosis isolates by using oligos MP1 and MT6x. The PCR product was purified as described above, and the sequence was determined directly by GATC.

Southern blot analysis. Digested genomic DNA was separated by electrophoresis on 1% agarose gels, transferred to positively charged nylon membranes (Roche Diagnostics), and treated according to the manufacturer's instructions. Probes for CaPAPx and CaMTL1a were amplified from C. albicans SCS314 genomic DNA by using primers MP1R and MTX6. All probes were labeled with [32P]dCTP by using a High Prime DNA Labeling Kit (Roche Diagnostics). For low-stringency conditions, hybridizations were carried out overnight at 60°C in PerfectHyb (Sigma). The blots were washed twice for 30 min in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)-0.1% sodium dodecyl sulfate (SDS) at 60°C (6). High-stringency hybridizations were carried out at 68°C, and the blots were washed once for 5 min at room temperature in 2× SSC-0.1% SDS and twice for 20 min at 68°C in 0.5× SSC-0.1% SDS. The blots were then exposed to X-ray film.

Analysis of RNA splicing. C. parapsilosis CLIB214 was cultured overnight in YPD medium, diluted to an A600 of 0.2 in 50 ml of fresh YPD medium, and incubated at 30°C for 4 h. Total RNA was isolated from 5 ml of culture by using a Ribopure kit (Ambion). DNase I treatment was carried out during the purification process. To generate cDNA, 3 μg of RNA and 291 ng of oligo dT-15 primer in a final volume of 15 μl was incubated at 70°C for 10 min and then chilled on ice. A 45-μl cocktail containing 0.6 mM deoxynucleoside triphosphates, 120 U of RNasin RNAse inhibitor (Promega), 600 U of Moloney murine leukemia virus reverse transcriptase (Promega), and 1× Promega reverse transcription-PCR (RT-PCR) buffer was then added. The mixture was incubated at

France. All other C. parapsilosis isolates were obtained from Frank Odds, Aberdeen, United Kingdom, and are described by Tavanti et al. (47).

Where indicated, a 200-bp fragment of the MTL1a region was amplified and the DNA sequence was determined.

TABLE 1. C. parapsilosis strains used in this study

| Isolate* | Geographic origin | Source | MTL1a sequenced*
|----------|------------------|--------|-----------------|
| CLIB214  | Puerto Rico      | Feces  | Yes             |
| 71701    | Aberdeen, United Kingdom | Unknown | Yes             |
| 73/107   | London, United Kingdom | Mouth | Yes             |
| 73/1058  | Belgium          | Nail   | Yes             |
| 39/1604  | Japan            | Unknown | Yes             |
| 81/042   | London, United Kingdom | Ear | Yes             |
| 81/041   | Mayo Clinic, Rochester, Minn. | Vagina |                 |
| 81/040   | London, United Kingdom | Toe space |                 |
| 73/037   | Leeds, United Kingdom | Vagina |                 |
| 81/253   | London, United Kingdom | Nail | Yes             |
| 74/046   | Leeds, United Kingdom | Aortic valve |                 |
| 90–137   | San Jose, Calif. | Orbital tissue | Yes             |

* All isolates, apart from the type strain, were obtained from Frank Odds, Aberdeen, United Kingdom, and are described by Tavanti et al. (47).

**Where indicated, a 200-bp fragment of the MTL1a region was amplified and the DNA sequence was determined.

TABLE 2. Oligonucleotides used in this study

| Oligo | Sequence |
|-------|----------|
| CP4   | 5′-CATTTCCGTAAAGCCTCTCAGATGTCTTTCGTC-3′ |
| CP2   | 5′-CCAAAATCTCAGCATACGTGATCGCAAAGCGTC-3′ |
| CP6   | 5′-CAGATGATGACTGCAATCCACCTTCAACTACAATGAC-3′ |
| CP3   | 5′-CGCACATTGTAACATCATAGGAAATATTATAGTTGGT-3′ |
| 3190_1 | 5′-TCTAGGGATGTTGACATGATAAATACGGGACCA-3′ |
| CCT7_1 | 5′-CGATGACATGCTACCGTCACAGCTTTC-3′ |
| MPR1  | 5′-GAGCTCTTGTGAAGAATTATTTCTCTGTA-3′ |
| MFX1  | 5′-AATCTGCTGTAATCAGATGACACAC-3′ |
| MFX2  | 5′-GCACACTGACAGATGTTTGTTGG-3′ |
| MFX3  | 5′-ATGGATCCTGCCGATTATTTAAAGCCT-3′ |
| MFX4  | 5′-TGGTATAAGTTTCCTCTTTGCTG-3′ |
| MTX5  | 5′-GGACTGTGCGTAATCACCACAG-3′ |
| MTX6  | 5′-TACCTTCTATTTTTTTGTTGCGG-3′ |
| MAR3  | 5′-GTGTCGATGAACAGATCCGACC-3′ |
| MAR4  | 5′-CGTTAGATTGTTGTTGGTGTG-3′ |
| MAR5  | 5′-GAATCTTATGATGCCATGTTGATG-3′ |
| MAR6  | 5′-GTAGGATGTGTTGTGTGTTCT-3′ |
37°C for 1 h, followed by 95°C for 2 min, and the resulting cDNA was held at 4°C. For each PCR, 5 μl of cDNA was used as a template in a 50-μl volume containing 0.25 mM deoxynucleoside triphosphates, 1/100 reaction buffer, 1.5 mM MgCl2, 25 pmol of each primer, and 2.5 U of Taq DNA polymerase (Invitrogen).

RNA that was not reverse transcribed was used as a control to test for the presence of genomic DNA. The same PCR amplification was also performed with genomic DNA as a template. The MTLa2 region was amplified by using oligos MTX1 and MTX2, and the MTLa1 region was amplified by using oligos MTX3, MTX4, MTX5, and MTX6. The MTX1/MTX2 and MTX3/MTX6 RT-PCR products were combined from five reactions, purified as described above, and the DNA sequence was determined directly by GATC.

RESULTS

Genome sequence survey of C. parapsilosis. We constructed a plasmid library with random genomic inserts of ca. 5 kb and sequenced both ends of ca. 8,000 clones. After removal of mtDNA and rRNA, 10.1 Mb of high-quality primary sequence data was obtained, corresponding to ca. 0.8x coverage of the haploid genome size. The sequences were assembled into 5,374 contigs. Putative ORFs were identified by using BLASTX searches of the C. albicans assembly 19 ORFs. We identified partial sequence of 3,898 orthologs in C. parapsilosis, corresponding to ca. 60% of the 6,419 nonoverlapping ORFs predicted in the haploid C. albicans genome (16). We also identified orthologs of seven additional ORFs found in other fungal genomes but which are not present in the C. albicans genome sequence (Table 3). The average protein sequence identity between C. parapsilosis and C. albicans is 59%. In contrast, the average identity between C. dubliniensis and C. albicans is 82% (based on a comparison of 5,511 C. albicans genes).

Since the cellular function of the majority of C. albicans genes is not yet determined, we inferred gene function in C. parapsilosis from the known functions of the S. cerevisiae orthologs. The “cellular role” divisions of the Yeast Proteome

FIG. 1. Identification of an MTLa idiomorph from C. parapsilosis. (A) The top line shows the structure of the MTLa sequence and surrounding region from C. albicans. The genes are drawn approximately to scale. The region indicated by a bracket is unique to MTLa, and the remaining sequences surround both MTLa and MTLa2 idiomorphs. The line below illustrates the corresponding structure in the orthologous CpMTLa2 idiomorph that was sequenced in full. The C. albicans genome annotation places an ORF (Orf19.3194) between CaYAP3 and CaGAP1. However, it is not conserved and is likely to be spurious and so is omitted in this diagram. Orf19.3192 corresponds to the N terminus of the Sti1 protein in S. cerevisiae. The sequence of the Cp19.3202 and Cp19.3192 orthologs at the ends of the fragment is not complete. Distances are shown in base pairs. The percent identity between proteins from C. parapsilosis (Cp) and C. albicans (Ca) or S. cerevisiae (Sc) is shown below. (B) An alignment of the MTLa2 proteins from C. parapsilosis (CpMTLa2) and C. albicans (CaMTLa2) was generated by using CLUSTAL W and is shown on the left-hand side. The intron locations are indicated with black triangles. The right-hand side shows an alignment between the C. albicans MTLa1 protein (CaMTLa1) and the hypothetical protein translation of the pseudogene from C. parapsilosis (CpMTLa1). Introns in the C. albicans sequence are indicated with black triangles, and a putative intron in C. parapsilosis is indicated with a gray triangle. Black circles highlight four stop codons in the C. parapsilosis sequence.
TABLE 3. Some C. parapsilosis genes not present in C. albicans

| Gene name* | Organism            | GenBank accession no. |
|------------|---------------------|-----------------------|
| DEHA0B13200g (capsid protein, virus L-A) | Debaryomyces hansenii | CAJ85523              |
| MG10213.4 (Gentisate 1,2-dioxyngease) | Magnaporthe grisea | EAA48150              |
| DEHA0E11440g (LMBR1-like membrane protein) | Debaryomyces hansenii | XP_459779              |
| FG04776.1 (monooxygenase) | Gibberella zeae | EAA75735              |
| SPAC11D3.03c | Schizosaccharomyces pombe | NP_592800              |
| DEHA0E1541g (esterase) | Debaryomyces hansenii | XP_459950              |
| KLL0A12045g | Kluyveromyces lactis | XP_451524              |

* The name of the closest ortholog is shown and, where possible, a function is inferred from sequences in other organisms and is shown in parentheses.

Database were used to categorize a total of 2,650 C. parapsilosis genes with identifiable S. cerevisiae orthologs. A similar analysis was performed with 7,677 ORFs from C. albicans derived from the haploid gene set (16; see also Table S5 in the supplemental material). We identified orthologs for 47% of the S. cerevisiae genes in C. parapsilosis and 51% in C. albicans. The category “conjugation with cellular fusion” included 94 genes from C. parapsilosis, several with known functions in mating. Although many have roles in other cellular processes, we extracted a list of 19 that function predominantly in mating (Table 4). These were augmented with orthologs of genes shown to be mating specific in C. albicans by Tsong et al. (48). The genes were placed into three classes: a-specific, α-specific, and genes more generally involved in mating. Although some may have several roles (such as Ram1 and Ram2, which are required for farnesylation of a-factor and Ras proteins in S. cerevisiae), others (such as Ste2, the α-factor receptor in S. cerevisiae) have no known function outside mating.

Identification of a putative C. parapsilosis MTL locus. In C. albicans, mating type is determined by the genes residing at the MTL locus (13, 48). There are two idiomorphs. MTLα contains the genes PAPα [poly(A) polymerase], OBPa (oxysterol-binding protein), PIKα (phosphatidyl inositol kinase), MTLα2 (a HMG-box protein), and MTLα1 (a homeodomain protein) (Fig. 1). MTLα contains three genes related to those in the MTLα locus (PAPα, OBPa, and PIKα), plus MTLα2 (a homeodomain protein) and MTLα1 (an α domain protein). The genes surrounding the MTL idiomorphs are the same on both chromosomes. Under certain conditions, cells with two copies of MTLα (α/α; or hemizygous for MTLα) can mate with cells with two copies of MTLα (α/α; or hemizygous for MTLα).

While analyzing the data from the sequence survey of C. parapsilosis, we noted that several reads matched many of the genes within and surrounding the MTLα locus. The sequence reads matching MTLα1, however, contained stop codons in the predicted ORF. To confirm this observation, we determined the sequence of the entire region orthologous to C. albicans MTLα (Fig. 1). This region contains orthologs of all of the genes found in the C. albicans MTLα locus, and the order and transcriptional orientations of the genes are conserved between the two species.

The C. parapsilosis MTLα1 locus is a pseudogene. Although the majority of the genes in the C. parapsilosis MTLα idiomorph are intact, CpMTLα1 appears to be a pseudogene, because there are four stop codons in the region corresponding

![](image)
to CaMTLa1 (Fig. 1B). Since C. parapsilosis CLIB214 is most likely diploid (7, 49), we considered the possibility that the genome might contain two copies of the MTLa1 idiomorph, one with a defective and one with a functional MTLa1 sequence. We therefore amplified a portion of the MTLa1 region from genomic DNA, derived from the conserved homeodomain and encompassing the last two stop codons. The sequence of the PCR product was determined directly. If two alleles are present we should obtain a mixed DNA sequence. However, only one sequence was detected, identical to the CpMTLa1 sequence (Fig. 1). We also amplified this fragment from seven other C. parapsilosis isolates (Table 1). All contained the same sequence, with no evidence of heterozygosity.

To confirm there is no copy of CpMTLa1 elsewhere in the genome, we performed a Southern blot, using the same 200-bp CpMTLa1 fragment as a probe. As shown in Fig. 2, only one fragment of the predicted size is detected under low-stringency conditions in the genomes of all C. parapsilosis isolates tested. Since C. albicans probes hybridize to C. parapsilosis DNA under these conditions (Fig. 4), it is highly likely that we would detect any related CpMTLa1 sequence present. Identical results were obtained at high stringencies (not shown).

The structure of CpMTLa1 as shown in Fig. 1 has one intron at a position equivalent to intron 2 of CaMTLa1, although the predicted 5’ splice site is GG, and there is no TACTAAC box (Fig. 3). The gene also has four internal stop codons. We considered another possible structure for this gene, one that incorporated an additional intron at a position similar to, but not identical to, intron 1 of the C. albicans gene (Fig. 1 and 3). The possible intron has GT and AG splice sites and a TAC TAAC lariat site and skips over two of the four predicted stop codons, but it also drastically shortens the predicted C. parapsilosis protein compared to C. albicans and other a1 proteins. To determine whether any splicing occurs, we compared the sizes of products obtained by RT-PCR from an RNA template to those generated by direct amplification of the genomic DNA. Primers were chosen that amplify the entire MTLa1 region, including the two possible introns (Fig. 3). As a control, we also amplified the MTLa2 region, which is predicted to contain an intron in both C. albicans and C. parapsilosis. As shown in Fig. 3, the RT-PCR product from CpMTLa2 is indeed smaller than the product amplified from genomic DNA. Sequence analysis confirmed that the predicted intron in MTLa2 is correctly spliced. To our knowledge, this is the first experimental demonstration of splicing of MTLa2 in Candida species. The RT-PCR products from CpMTLa1, however, are identical in size to those derived from genomic DNA. The sequence of the largest RT-PCR product was determined, confirming that it is generated from the CpMTLa1 sequence and that no splicing occurs. We are thus unsure whether CpMTLa1
is a pseudogene of a single-intron gene that has gained four stop codons or a pseudogene of a two-intron gene that has gained two stop codons, but in either case it is undoubtedly a pseudogene.

**Structure of the C. parapsilosis MTL locus.** Our genome sequencing data yielded several random genomic clones that originated from MTLα, but none that matched MTLα. To look for MTLα sequences in this and other isolates, we first used overlapping clone data to extrapolate the structure of the region surrounding the CpMTL locus (Fig. 4). The order of the surrounding genes is identical to the MTL loci in C. albicans (16). We used four primer combinations to amplify the region surrounding the MTL locus from C. parapsilosis CLIB214, yielding fragments varying in size from 14.2 to more than 21 kb (Fig. 4). Primer CP6 lies within the region for which the entire sequence was determined (Fig. 1), and the remaining primers were derived from the sequences of overlapping clones. If two idiomorphs are present we should amplify two fragments. However, as shown in Fig. 4B, restriction analysis shows that only one fragment is present. The restriction pattern is also identical for eight isolates amplified by using the primer pair CP6 and CCT7_1 (Fig. 4C), and for 14 isolates amplified by using CP6 and CP3 (not shown), indicating that these PCR products also contain MTLα idiomorphs.

To look for evidence of CpMTLA sequences elsewhere in the genome, we performed low-stringency Southern blot analysis. We first used an 856-bp fragment derived from CaPAPα, which is found in the MTLα idiomorph of C. albicans (13), C. dubliniensis (38), and C. tropicalis (http://www.broad.mit.edu). We reasoned that at low stringency, this probe should hybridize to the related gene CpPAPα, which we have shown is present in the CpMTLα idiomorph, and to any ortholog of PAPα present. As shown in Fig. 5, CaPAPα hybridizes to DNA fragments of 2.7 kb (BglII digest) and 2.9 kb (BamHI/PstI digest). These correspond to the predicted size of CpPAPα, which we confirmed in other blots (not shown). However, it hybridizes at higher intensities to a larger fragment in both digests (>8 kb), suggesting the C. parapsilosis genome does contain an ortholog of CaPAPα. A 145-bp fragment derived from the conserved domain of CaMTLα1 also hybridized to sequences in C. parapsilosis (Fig. 5B). Both probes detected large fragments in all of the digests tested (not shown), and we were unable to determine whether the sequences are contiguous. It is therefore likely that C. parapsilosis does contain MTLα sequences, but the PCR results suggest that they are not in the same chromosomal context as CpMTLα, even though the order of the genes flanking CpMTLα is very similar to those flanking the MTL loci of C. albicans, C. dubliniensis, and C. tropicalis (Fig. 4).

**DISCUSSION**

The results of the genome sequence survey of C. parapsilosis yielded partial DNA sequence information for more than 3,900 potential ORFs. The amino acid sequence identity between C. parapsilosis proteins and C. albicans proteins and between C. dubliniensis proteins and C. albicans proteins both ranged from ca. 20 to 100%. The most similar sequences included ribosomal proteins, actin, and polyubiquitin (see Table S6 in the supplemental material). Proteins with <25% conservation between

C. parapsilosis and C. albicans are mostly uncharacterized, but include AXL1, a putative a-pheromone maturase (see Table S7 in the supplemental material) (1). The average sequence identity between C. parapsilosis and C. albicans proteins, however, is 59%, and between C. dubliniensis and C. albicans it is 82%.

To identify genes that may be unique to C. parapsilosis, we
extracted sequences with no similarity to the \textit{C. albicans} genome and looked for orthologs in other species. We identified seven genes all present in other fungal genomes (Table 3). One is strongly similar to the major coat protein of the \textit{S. cerevisiae} L-A virus, suggesting that \textit{C. parapsilosis} may harbor double-stranded RNA viruses or that its genome contains genes derived from such a virus. In other yeasts these viruses encode toxins, some of which have anti-\textit{Candida} activities (42, 43, 50). It will therefore be interesting to look for similar activities in \textit{C. parapsilosis}.

\textbf{Disruption of the mating locus in \textit{C. parapsilosis}.} In \textit{C. albicans}, the ability to mate is closely associated with white-opaque phenotype switching. In heterozygous diploid cells containing both \textit{a} and \textit{\alpha} idiomorphs, \textit{a1} and \textit{\alpha2} proteins heterodimerize to generate a negative regulator. As a result, white-opaque switching is repressed and mating-specific functions are switched off (27, 31, 32, 48). White-opaque switching occurs in \textit{C. albicans} cells that have only one functional \textit{MTL} idiomorph because they are homozygous or hemizygous (24, 27). In opaque cells that contain only the \textit{MTL\alpha} idiomorph, the \textit{\alpha2} protein is a positive regulator of mating functions. Similarly, \textit{\alpha1} positively regulates mating in opaque cells containing only the \textit{MTL\alpha} idiomorph. The closely related species \textit{C. dubliniensis} also undergoes white-opaque switching and mates in a similar fashion (38). Phenotypic switching does occur in \textit{C. parapsilosis} (9, 23, 28), but we have not observed anything directly comparable to white-opaque phenotypes. We have, however, shown that the type strain of \textit{C. parapsilosis} carries a defective \textit{MTL\alpha1} allele, with up to four internal stop codons in the presumptive ORF.

The final two stop codons in \textit{CpMTL\alpha1} from \textit{C. parapsilosis} CLIB214 are also present in seven other randomly selected isolates, and PCR amplification of genomic DNA and Southern blot analysis indicates that there is no functional copy of \textit{MTL\alpha1} elsewhere. It is therefore extremely unlikely that \textit{C. parapsilosis} can generate a heterodimeric \textit{a1/\alpha2} protein. The \textit{MTL\alpha2} gene, however, is transcribed and spliced, and so should be capable of inducing \textit{a}-specific mating type functions.

If mating occurs in \textit{C. parapsilosis} in a manner that is in any way similar to that in \textit{C. albicans} or \textit{C. dubliniensis}, we would expect to find an \textit{MTL\alpha} idiomorph in an analogous position to \textit{MTL\alpha}. Our sequence survey resulted in 0.8× coverage of the genome, with several matches to \textit{MTL\alpha}-specific genes such as \textit{PIKa}, \textit{MTL\alpha1}, and \textit{MTL\alpha2}. No similarities to \textit{MTL\alpha} genes were detected. To look for a second idiomorph, we first compared the structures of the \textit{MTL} loci from related \textit{Candida} species. In \textit{C. albicans}, the region unique to \textit{MTL\alpha} is 8,742 bp, and the unique \textit{MTL\alpha} sequence is 8,861 bp (13). The equivalent unique regions in \textit{C. dubliniensis} are 8,728 bp (\textit{CdMTL\alpha}) and 8,845 bp (\textit{CdMTL\alpha1}) (http://www.sanger.ac.uk/Projects/C_dubliniensis/), and in \textit{C. tropicalis} they are 8,473 bp (\textit{CtMTL\alpha}) and 10,337 bp (\textit{CtMTL\alpha1}) (http://www.broad.mit.edu). The gene order within the equivalent idiomorphs is identical (Fig. 4). It is not possible to determine the order of all of the genes surrounding the \textit{MTL} loci, since the assemblies for \textit{C. dubliniensis} and \textit{C. tropicalis} are not complete. However, the synteny is very similar for all of the loci for which sequence is currently available (Fig. 4), although there does appear to be a break upstream from \textit{STI1} in \textit{C. dubliniensis} \textit{MTL\alpha}. We therefore amplified regions surrounding \textit{CpMTL} by using primers derived from the sequence of the unique genes that lie well outside the borders of the \textit{MTL} locus in related \textit{Candida} species (Fig. 4A). We amplified only one DNA fragment, corresponding to \textit{CpMTL\alpha1}, in all isolates tested (Fig. 4B and C). Since there is very little sequence variability in \textit{C. parapsilosis} isolates (46) and we used four primer combinations, it is highly
unlikely that our failure to amplify an MTLα-containing fragment was caused by polymorphisms in the primer sequences. It therefore appears that the chromosomal location that corresponds to MTL in other Candida species contains only MTLα in C. parapsilosis. However, our Southern blot analysis does suggest that there are MTLα-like sequences somewhere in the C. parapsilosis genome, since we identified fragments that hybridize to CarpAα and CaMTLa1 (Fig. 5). We conclude that MTLαs in C. parapsilosis has been rearranged somehow, breaking its colinearity with the genomic context of CpMTLa1 and the MTL loci in other Candida species. It is difficult to envisage how C. parapsilosis strains could become mating competent through homoyzogosity or hemizygosity at the MTL genes, when these genes are not at allelic locations in the genome. It is also unlikely that the MTLα sequences represent a silent mating cassette involved in mating type switching, since our analysis has shown that this mechanism evolved relatively recently and is restricted to the Saccharomyces sensu stricto group and their close relatives (5). The most likely scenario is that the MTLα locus has degenerated after rearrangement and has become inactivated, although further sequence analysis is required to support this hypothesis. Despite the overall similarities, however, it is very unlikely that C. parapsilosis can mate in the same way as C. albicans (15) or C. dubliniensis (38).

Despite the defects in the MTL sequences, we did identify many genes that are predominantly required for mating in C. albicans and S. cerevisiae (Table 4). These include orthologs of STE11, required for processing of α-factor (17), and STE2, the α-factor receptor in S. cerevisiae (4). We also identified several genes that are probably α-specific, including orthologs of STE6, RAM2, Orf19.4688, Orf19.2825, and RBT1 whose expression in C. albicans requires both MTLα2 and exposure to α-factor (48). STE6 is known to be required for mating in C. albicans (30). The fact that many components of the pathway are apparently intact suggests that C. parapsilosis has only recently undergone a rearrangement at the MTL locus. However, this contrasts with the observation that the MTLα1 sequence contains four stop codons, which were presumably acquired in four separate mutational events.

ACKNOWLEDGMENTS

This study was supported by Science Foundation Ireland. We are grateful to Jozef Nosek for providing access to the C. parapsilosis mitochondrial genome prior to publication and to Frank Odds for providing the strain collection.

REFERENCES

1. Adams, N., K. Blundell, M. N. Ashby, and C. Boone. 1995. Role of yeast insulin-degrading enzyme homologs in pheromone processing and bud site selection. Science 270:464–467.
2. Baillie, G. S., and L. J. Douglas. 1998. Effect of growth rate on resistance of Candida albicans biofilms to antifungal agents. Antimicrob. Agents Chemother. 42:1900–1905.
3. Brockert, P. J., S. A. Lachke, T. Srikantha, C. Pujol, R. Galask, and D. R. Soll. 2003. Phenotypic switching and mating type switching of C. parapsilosis in United States hospitals: a three-year analysis. Clin. Infect. Dis. 29:239–244.
4. Enger, L., S. Joly, C. Pujol, P. Simonson, M. Pfaffer, and D. R. Soll. 2001. Characterization and cloning of a complex DNA fingerprinting probe for Candida parapsilosis. J. Clin. Microbiol. 39:655–669.
5. Fundygda, R. E., R. J. Kukendall, W. Lee-Yang, and T. J. Lott. 2004. Evidence for aneuploidy and recombination in the human commensal yeast Candida parapsilosis. Infect. Genet. Evol. 4:37–43.
6. Hajhej, R. A., A. N. Sofair, L. H. Harrison, G. M. Lyon, B. A. Arthington-Hull, C. M., R. M. Raisner, and A. D. Johnson. 2000. Evidence for mating of the “asexual” yeast Candida parapsilosis in a mammalian host. Science 293:307–310.
7. Hull, C. M., R. M. Raisner, and A. D. Johnson. 2000. Evidence for mating of the “asexual” yeast Candida parapsilosis in a mammalian host. Science 293:307–310.
8. Hulit, K. Blundell, M. N. Ashby, and C. Boone. 1995. Role of yeast insulin-degrading enzyme homologs in pheromone processing and bud site selection. Science 270:464–467.
9. Butler, G., C. Kenny, A. Fagan, C. Kurischko, C. Gaillardin, and K. H. Wolfe. 2004. Evolution of the MAT locus and its Ho endonuclease in yeast species. Proc. Natl. Acad. Sci. USA 101:1632–1637.
10. Cho, Y., Y. L. Qiu, P. Kuhiman, and J. D. Palmer. 1998. Explosive invasion of plant mitochondria by a group I intron. Proc. Natl. Acad. Sci. USA 95:14244–14249.
11. Dvorak, J. A., W. L. Whelan, J. P. McDaniel, C. C. Gibson, and K. J. Kwon-Chung. 1987. Flow cytometric analysis of the DNA synthetic cycle of Candida species. Infect. Immun. 55:1490–1497.
12. Edmond, M. B., S. E. Wallace, D. K. McClish, M. A. Pfaffer, R. N. Jones, and R. P. Wenzel. 1999. Nosocomial bloodstream infections in United States hospitals: a three-year analysis. Clin. Infect. Dis. 29:239–244.
13. Enger, L., S. Joly, C. Pujol, P. Simonson, M. Pfaffer, and D. R. Soll. 2001. Characterization and cloning of a complex DNA fingerprinting probe for Candida parapsilosis. J. Clin. Microbiol. 39:655–669.
14. Fundygda, R. E., R. J. Kukendall, W. Lee-Yang, and T. J. Lott. 2004. Evidence for aneuploidy and recombination in the human commensal yeast Candida parapsilosis. Infect. Genet. Evol. 4:37–43.
15. Hajhej, R. A., A. N. Sofair, L. H. Harrison, G. M. Lyon, B. A. Arthington-Hull, C. M., R. M. Raisner, and A. D. Johnson. 2000. Evidence for mating of the “asexual” yeast Candida parapsilosis in a mammalian host. Science 293:307–310.
albicans is controlled by mating-type locus homeodomain proteins and allows efficient mating. Cell 110:293–302.

33. Nosek, J., M. Novotna, Z. Hlavatovicova, D. W. Ussery, M. Pfaller, and D. J. Diekema. 2001. Complete DNA sequence of the linear mitochondrial genome of the pathogenic yeast Candida parapsilosis. Mol. Genet. Genomics.

35. Pagano, L., A. Antinori, A. Ammassari, L. Mele, A. Nosari, T. Srikantha, J. B. Radke, J. Geiger, S. J. Kee, M. G. Shin, S. H. Kim, J. B. Radke, R. T. Wiblin, and K. J. Daniels. 2004. Evidence from comparative genomics for a complete sexual cycle in the ‘asexual’ pathogenic yeast Candida glabrata. Genome Biol. 4:R10.