Multiple gene genealogical analyses suggest divergence and recent clonal dispersal in the opportunistic human pathogen *Candida guilliermondii*

Lisa Lan and Jianping Xu

Correspondence
Jianping Xu
jpxu@mcmaster.ca

Department of Biology, McMaster University, 1280 Main St West, Hamilton, ON L8S 4K1, Canada

Candida guilliermondii is a haploid opportunistic pathogen accounting for about 2 % of human blood yeast infections. Recent analyses using multilocus enzyme electrophoresis and karyotyping suggest that strains from human sources traditionally designated *C. guilliermondii* in fact include at least two species, *C. guilliermondii* and *Candida fermentati*. However, the patterns of molecular variation within and between these two species remain largely unknown. In this study, DNA fragments were sequenced from five genes for each of 37 strains collected from Canada, China, the Philippines and Tanzania. The analyses identified significant sequence differences between *C. guilliermondii* and *C. fermentati*. The five gene genealogies showed no apparent incongruence, suggesting a predominantly clonal reproductive structure for both species in nature. Indeed, two large clones of *C. guilliermondii* were identified, with one from Ontario, Canada, and the other from China. Interestingly, the results indicate that strains currently designated *C. guilliermondii* may contain additional divergent lineages. On the practical side, the results revealed several diagnostic molecular markers that can be used in clinical microbiology laboratories to distinguish *C. guilliermondii* and *C. fermentati*. The multiple gene genealogical analyses conducted here revealed significant divergence and clonal dispersal in this important pathogenic yeast complex.

Received 24 October 2005

Revised 17 January 2006

Accepted 18 January 2006

INTRODUCTION

Emerging infectious diseases are increasingly recognized as significant threats to human health and social and economic wellbeing. Most emerging infections are believed to result from anthropogenic activities such as increased contact between previously disjunct human and wildlife populations, the disturbance of human and animal ecological niches, and the rapid dispersal of infectious agents accompanying human travel. A prominent feature of emerging human infectious agents is the widespread geographical distribution of clones and genotypes. For example, in the model human fungal pathogens *Candida albicans* and *Cryptococcus neoformans*, strains with similar or identical multilocus genotypes are commonly found among diverse geographical locations in South and North America, Australia, Eurasia and Africa (Kidd *et al.*, 2005; McEwen *et al.*, 2000; Xu, 2004a; Xu *et al.*, 1999b, c, 2000b). While most of our attention has focused on model pathogens, the population biology of other opportunistic fungal pathogens such as *Candida guilliermondii* remains poorly understood.

C. guilliermondii is a haploid ascomycete yeast (Doi *et al.*, 1992; Imshenetskii *et al.*, 1979). Like most other species in the genus *Candida*, *C. guilliermondii* is a common constituent of the normal human microflora (Kam & Xu, 2002; Odds, 1988; Tey *et al.*, 2003; Xu & Mitchell, 2003a). However, when the immune system of the host is compromised due to organ transplantation, chemotherapy, or infections by other agents such as the human immunodeficiency virus, *C. guilliermondii* and other *Candida* species can cause severe, sometimes life-threatening infections (Xu *et al.*, 1999a). Whereas *Candida albicans* is the most commonly isolated pathogenic yeast in humans, other species such as *C. guilliermondii* are commonly found (Dos Santos & Soares, 2005; Krcmery & Barnes, 2002; Odds, 1988; Sandven, 2000; Tey *et al.*, 2003). Globally, *C. guilliermondii* accounts for about 2 % of all candidaemia (Krcmery & Barnes 2002; Sandven, 2000; Tey *et al.*, 2003). However, in certain geographical areas such as Italy, India and Brazil, it may account for over 10 % of all candidaemia (Sandven, 2000).

Unlike the more common *C. albicans*, whose primary ecological niche is on or in mammals and for which a complete sexual life cycle has not been found, *C. guilliermondii* can be found in a variety of ecological niches and has an identifiable sexual life cycle in the laboratory (Odds, 1988; Barnett *et al.*,

Abbreviations: I_A , index of association; K2P, Kimura two-parameter; MP, maximum parsimony; PH, partition homogeneity; T-PTP, topology-dependent permutation tail probability.

2000). However, whether such distinct lifestyles contribute to different population genetic patterns remains unknown. Before the mid-1990s, the identification of human pathogenic (as well as non-pathogenic) fungi was mostly based on morphological and physiological features. The use of molecular markers in the last decade has uncovered several previously unknown species such as *Candida dubliniensis*, a close relative of *C. albicans* (Sullivan *et al.*, 1995), and two close relatives of *C. parapsilosis*, *C. orthopsilosis* and *C. metapsilosis* (Tavanti *et al.*, 2005). In clinical samples of *C. guilliermondii*, multilocus enzyme electrophoresis (MLEE) and karyotyping using pulsed-field gel electrophoresis (PFGE) revealed that this species in fact contained two different species, *C. fermentati* and *C. guilliermondii* (Bai *et al.*, 2000; San Millan *et al.*, 1997). However, readily applicable diagnostic markers to distinguish these two species remain elusive. Furthermore, the patterns of DNA sequence variation within and between these two closely related species remain unknown.

The objective of this study was to use the multiple gene genealogy (also called multilocus sequence typing or MLST; Maiden *et al.*, 1998) approach to understand the evolution and population genetic structure of *C. guilliermondii* and *C. fermentati*. Gene genealogical analyses have been used to address a variety of ecological and evolutionary genetic issues, including the rates of dispersal, hybridization and divergence, and the role of clonality and recombination in the structure of fungal populations (Xu, 2006; Xu *et al.*, 2002, 2005; Xu & Mitchell, 2003b). Given a group of clonally reproducing organisms, the genealogies of multiple genes are expected to be the same, since evolutionary changes arise mainly through independent mutations. In contrast, in a group of organisms in which sexual recombination has occurred to some degree, the gene genealogies may be expected to differ, since sexual mating and meiosis result in the shuffling of genes from different parts of a genome (Xu, 2004b).

This study was designed to address the following specific questions. First, how much divergence is there between *C. guilliermondii* and *C. fermentati* at the DNA sequence level? Will the qualitative differences between the two species as revealed by MLEE and karyotyping be substantiated at the DNA sequence level? Second, are genealogies from different genes congruent? Congruent gene genealogies suggest a predominantly clonal mode of reproduction in natural populations of the species. Because *C. guilliermondii* is capable of sexual reproduction in the laboratory, we hypothesize that there should be evidence for gene genealogy incongruence and recombination in natural populations of *C. guilliermondii*. Third, what is the geographical pattern of molecular variation? Specifically, are strains from the same area more similar to each other than to those from other areas? And fourth, are there widespread clones or clonal lineages in *C. guilliermondii* as has been found for several other human fungal pathogens (Pujol *et al.*, 2005; Xu, 2004a)? To address these questions, we analysed a collection

of 37 strains using DNA sequences from fragments of five genes.

METHODS

Yeast isolates. Thirty-seven isolates were obtained from diverse geographical areas in four countries. Three strains were isolated from two healthy hosts in an aboriginal community at the First Nations Reserve just west of Hamilton, Ontario, Canada. Twelve strains were obtained from the regional Clinical Microbiology Laboratory in Toronto, Ontario. All 12 strains were from unrelated patients with yeast infections by *C. guilliermondii*. Six strains were collected from separate patients in the Philippines. Two strains were isolated from the oral cavities of patients infected with HIV in Tanzania. Fourteen strains were collected from three provinces in China: eight strains from a small farming village (Shitou Village) in Jiangxi Province in southeast China; one strain from Xi'an, the capital of Shaanxi Province in central China; and five strains from Chengdu, the capital of Sichuan Province in southwest China.

These 37 strains were initially identified as *C. guilliermondii* based on standard clinical mycological procedures using growth phenotypes on agar plates, microscopic features and API32 substrate utilization profiles (Warren & Hazen, 1999; Kam & Xu, 2002). Their identifications were confirmed and further distinctions were made using sequence analysis of the internal transcribed spacer (ITS) regions 1 and 2 using the PCR primers ITS1 and ITS4. These 37 strains were found to belong to two different species, 32 being *C. guilliermondii* and 5 being *C. fermentati* (see below). Details of these strains and their species designations are presented in Table 1. Except for strains LM85 and LM86, which were obtained from the same host but different body sites, all strains were isolated from a distinct host/source (Table 1).

DNA manipulation. High-molecular-mass genomic DNA was isolated using a previously described technique (Xu *et al.*, 2000a). Fragments of five genes were studied: *ACT1* (actin 1 gene), *EF1* (translation elongation factor 1), *ITS* of the rRNA gene cluster, *RIBO* (riboflavin synthetase) and *TOPOII* (DNA topoisomerase II). These genes were chosen because they were conserved, housekeeping genes, to ensure successful amplification and sequencing of gene fragments from all strains. The chromosomal locations, primer sequences and original GenBank accession numbers for these genes in *C. guilliermondii* are given in Table 2. These GenBank sequences were used to design specific primers for the amplification and sequencing of each gene from each of the 37 strains. The primers and the expected PCR product size for all five gene fragments are presented in Table 2. PCR amplification was performed according to standard protocols (Xu *et al.*, 2002, 2000b; Xu & Mitchell, 2003b). PCR products were cleaned using the microCLEAN purification kit (Qiagen) and sequenced using both the forward and reverse primers using an ABI3100 automated DNA sequencer (MoBix Laboratory, McMaster University). For each of the five gene fragments, strains that showed DNA sequences differing by only one or a few nucleotides from the dominant alleles were sequenced again to confirm the observed base substitutions. In all cases, the second round of sequencing produced sequences identical to those obtained in the first round.

Data analyses. Phylogenetic analyses of the five individual gene fragments as well as the combined data were performed using the PAUP* software (Swofford, 2004). The most parsimonious trees were obtained by heuristic searches based on 500 random sequence additions. Mid-point rooting was used for all phylogenetic trees.

To test for clonality and recombination in samples of the two species, we implemented three tests. In the first, we calculated the overall index of association (I_A). In this test, the observed data were compared

Table 1. Strains of *Candida guilliermondii* and *Candida fermentati* used in this study

Isolate ID	Source*	Geographical origin†	Multilocus genotype‡					Year isolated
<i>C. guilliermondii</i>								
LM8	HH, fingernail	First Nations Res., ON, CA	1	1	1	1	1	2002
TOR1	Patient	Toronto, ON, CA	1	1	1	1	1	2002
TOR2	Patient	Toronto, ON, CA	1	1	1	1	1	2003
TOR3	Patient	Toronto, ON, CA	1	1	1	1	1	2003
TOR4	Patient	Toronto, ON, CA	1	1	1	1	1	2004
TOR5	Patient	Toronto, ON, CA	1	1	1	1	1	2004
TOR6	Patient	Toronto, ON, CA.	1	1	1	1	1	2004
TOR7	Patient	Toronto, ON, CA	1	1	1	1	1	2003
TOR8	Patient	Toronto, ON, CA	1	1	1	1	1	2003
TOR9	Patient	Toronto, ON, CA	1	1	1	1	1	2004
TOR10	Patient	Toronto, ON, CA	1	1	1	1	1	2002
TOR12	Patient	Toronto, ON, CA	1	1	1	1	1	2002
TOR13	Patient	Toronto, ON, CA	1	1	2	1	1	2004
PH21	Patient, bone marrow	Philippines	1	1	1	2	1	1997
PH144	Patient, blood	Philippines	2	2	3	3	2	1996
PH145	Patient, peritoneal fluid	Philippines	2	3	3	4	3	1997
PH146	Patient, eyelid discharge	Philippines	1	1	1	2	1	1996
PH390	Patient, blood	Philippines	1	1	1	2	1	1996
ZG15	HH, oral cavity	Jiangxi, PRC	1	1	3	2	1	1997
ZG38	HH, oral cavity	Jiangxi, PRC	1	1	3	2	1	1997
ZG57	HH, oral cavity	Jiangxi, PRC	1	1	3	2	4	1997
ZG64	HH, oral cavity	Jiangxi, PRC	1	1	3	2	1	1997
ZG70	HH, oral cavity	Jiangxi, PRC	1	1	3	5	1	1997
ZG87	HH, oral cavity	Jiangxi, PRC	1	1	3	2	1	1997
ZG98	HH, oral cavity	Jiangxi, PRC	1	1	3	2	1	1997
ZG100	HH, oral cavity	Jiangxi, PRC	1	1	3	2	1	1997
ZG233	HH, oral cavity	Shaanxi, PRC	1	1	3	2	1	1997
ZG247	HH, oral cavity	Sichuan, PRC	1	1	3	2	1	1997
ZG248	HH, oral cavity	Sichuan, PRC	1	1	4	2	1	1997
ZG251	HH, oral cavity	Sichuan, PRC	1	1	3	2	1	1997
ZG258	HH, oral cavity	Sichuan, PRC	1	1	3	2	1	1997
ZG379	Bird dropping	Sichuan, PRC	1	1	3	6	1	1997
<i>C. fermentati</i>								
LM85	HH, mouth	First Nations Res., ON, CA	5	5	5	7	5	2002
LM86	HH, fingernail	First Nations Res., ON, CA	5	5	5	7	5	2002
PH20	Patient, foot scraping	Philippines	3	5	5	8	6	1997
TZ125	Patient, oral cavity	Tanzania	4	4	5	9	7	1998
TZ144	Patient, oral cavity	Tanzania	5	5	5	8	7	1998

*HH, healthy host.

†ON, Ontario; CA, Canada; PRC, People's Republic of China.

‡Multilocus genotype: alleles for genes (from left to right) *ACT1*, *EFI*, *ITS*, *RIBO* and *TOPOII* respectively.

against the null hypothesis that alleles from different loci were randomly associating with each other (i.e. a recombining population structure). A clonal population structure would be suggested if the null hypothesis were rejected; i.e. if alleles showed significant association with each other (Xu, 2005). In the second test, the proportion of pairwise loci that were phylogenetically incompatible was calculated. A phylogenetic incompatibility occurs between two loci with two alleles each when all four possible genotypes are found in the population. To perform the above two tests, each unique haplotype of each gene was counted as a unique allele and the data were then imported to the program MULTILOCUS version 1.3 (Agapow & Burt, 2001) for analyses.

The underlying assumptions, formulae, and the inferences of statistical significance of these two tests can be found on the program homepage (Agapow & Burt, 2001). In the third test, the congruence of gene genealogies was assessed using the partition homogeneity (PH) test (Farris *et al.*, 1995). In the PH test, the null hypothesis is congruence and clonality, the opposite of that for the index of association test (Xu, 2005). These three complementary tests were applied to samples of both *C. guilliermondii* and *C. fermentati* individually.

To test whether there were significant phylogenetic patterns based on geographical origins, we used the topology-dependent permutation tail

Table 2. Gene and primer information for the five genes analysed in this study

Gene	Chromosome/ supercontig no.*	Primer sequences (5'–3'; F, forward; R, reverse)	GenBank accession no.	Length of predicted PCR product (bp)
Actin 1 (<i>ACT1</i>)	7	F: ATCGTCGGTAGACCAAGACAT R: TTCCAATCCCAAGTCGGAA	AJ508528	711
Elongation factor 1 (<i>EF1</i>)	3	F: TCAACCTATCGACGAGGAAGT R: TCACCAGAGAAACCGAAAGA	AY497662	579
Gene region encoding partial 18S rRNA, ITS1, 5·8S rRNA, ITS2 and partial 26S rRNA (<i>ITS</i>)	5	F: TCCGTAGGTGAACCTGCGG R: TCCTCCGCTTATTGATATGC	AB054109	585
Riboflavin synthetase (<i>RIBO</i>)	1	F: ACAAGTTGGTCGAGGTGGTC R: CCTGGGTTCCCAAGTAGTCA	AF459790	597
DNA topoisomerase II (<i>TOPOII</i>)	5	F: AGCTTTATCAGCGTGGAAA R: GGCTGCTTGTTGGTCTCTTC	Z74991	703

*Location of the five genes on the chromosomes/supercontigs. The information was obtained from BLAST searches against the recent assembly of a 10× coverage sequencing of the *C. guilliermondii* genome by the Broad Institute (http://www.broad.mit.edu/annotation/fungi/candida_guilliermondii/background.html).

probability (T-PTP) test (Faith, 1991). This test compares the lengths of maximum-parsimony (MP) trees with and without the monophyletic constraint defined by the described trait, geography. If the constrained trees are significantly longer than the unconstrained MP tree, the results suggest a lack of phylogenetic pattern based on the specific trait. The statistical significance of the T-PTP test was derived from permutation of the sequence data under the assumption of non-monophyly to generate a null distribution of tree lengths. Statistical support for non-monophyly is achieved when over 95 % of all permuted datasets have tree lengths shorter than the MP tree generated with the constraint of monophyly (Faith, 1991). The T-PTP test was implemented in PAUP* (Swofford, 2004).

Based on the derived phylogenetic pattern, the existence of potentially divergent lineages within each of the two species was further analysed using evolutionary distances within and between groups of strains. In this study, the common Kimura two-parameter (K2P) distance implemented in PAUP* was used. The K2P model treats transitions and transversions differently and uses observed empirical substitution patterns to derive the optimum weighing schemes among various types of transitions and transversions (Swofford, 2004).

RESULTS

We successfully obtained DNA sequences from fragments of all five genes for all 37 strains. For each of the 185 (37 × 5) DNA sequences, we observed no evidence of heterozygosity (i.e. double peaks in the sequencing chromatograms). This observation is consistent with the haploid nature of the *C. guilliermondii* and *C. fermentati* genomes. In the following analyses, we therefore treat the strains as haploid. The composite multilocus genotype for each strain is presented in Table 1. The patterns of DNA sequence variation are described below.

Significant molecular divergence between *C. guilliermondii* and *C. fermentati*

The BLASTn analyses of the *ITS* gene sequences identified that the 37 strains analysed here belonged to two different species. Thirty-two strains had *ITS* sequences 99·7–100 % identical to that of the type sequence of *C. guilliermondii*, and five strains matched 100 % of the type sequence of *C. fermentati*. Within the ITS1 and ITS2 regions, four polymorphic nucleotide sites distinguish these two species (Fig. 1).

The other four gene fragments analysed here also showed consistent divergence between *C. guilliermondii* and *C. fermentati* (Fig. 1, Table 3). These results contrast with those from traditional identification methods using morphological characteristics and physiological features, which failed to distinguish strains of these two species. Among the five gene fragments, *RIBO* showed the greatest divergence between the two species, followed by *TOPOII*, *EF1*, *ACT1* and the least divergent *ITS* gene. The mean K2P distance between pairs of strains for *RIBO* (0·1413) was about twice that of *TOPOII* and almost 20 times that of the *ITS* gene fragment (Table 3). When all five genes are considered together, the mean K2P distance between *C. guilliermondii* and *C. fermentati* is 0·0511. This divergence is over 20 times the mean divergence between strains within both species. The results here thus support independent species status for *C. guilliermondii* and *C. fermentati*.

Limited nucleotide variation within populations of *C. guilliermondii* and *C. fermentati*

Except the *ITS* gene fragment, which showed no variation among the five strains of *C. fermentati*, all other gene

fragments analysed here showed various degrees of nucleotide variation within populations of these two species. However, compared to the extensive variation between the two species, the variation within samples of both species was very limited. In addition, the rank orders among the five genes in the degree of nucleotide variation within both species were different from that between the two species. Specifically, within *C. guilliermondii*, the most divergent was *RIBO*, followed by *EF1*, *TOPOII*, *ITS* and *ACT1*. Within *C. fermentati*, the most divergent was *TOPOII*, followed by *RIBO*, *ACT1*, *EF1* and *ITS*. These results indicated that the most divergent genes within these two species were not necessarily the most divergent between the species and vice versa.

The number of haplotypes among genes also varied in populations of the two species and was not directly correlated with the degree of nucleotide variation within or between the two examined species. Specifically, in *C. guilliermondii*, two haplotypes were found for *ACT1*, three for *EF1*, four each for *TOPOII* and *ITS*, and six for *RIBO* (Table 1). In *C. fermentati*, one haplotype was found for *ITS*, two for *EF1*, and three each for *ACT1*, *RIBO* and *TOPOII* (Table 1). The combined nucleotide data identified a total of 10 multilocus genotypes among the 32 strains of *C. guilliermondii* and 4 multilocus genotypes among the five strains of *C. fermentati*.

Evidence for clonal reproduction in *C. guilliermondii*

Several lines of evidence suggest that clonal reproduction is the main mode of reproduction in natural populations of *C. guilliermondii*. First, we observed that several multilocus genotypes were shared by multiple strains, all from unrelated hosts. Specifically, among 10 multilocus genotypes identified for the 32 strains of *C. guilliermondii*, three were shared by a total of 25 strains: one genotype was shared by three strains from the Philippines, one by 12 strains from Ontario, Canada, and the third by 10 strains from diverse geographical locations in China (Table 1, Fig. 2). The second line of evidence came from gene genealogical comparisons. The PH test revealed that genealogies from all five genes are congruent among each other in the sample of *C. guilliermondii* ($P=0.81$). The third line of evidence came from linkage disequilibrium analyses. The alleles from the five analysed loci showed no evidence of random association ($P<0.001$, Table 4). Indeed, we found no evidence for phylogenetic incompatibility in this sample of *C. guilliermondii* ($P<0.001$, Table 4).

Similar evidence exists for clonal reproduction in the natural population of *C. fermentati*. The PH test identified that genealogies from the five genes are statistically congruent among each other in this sample of *C. fermentati* ($P=0.48$). No evidence for phylogenetic incompatibility was observed. Furthermore, despite the small sample size, linkage disequilibrium analysis rejected the null hypothesis of random mating (Table 4). Among the five strains analysed here, two shared identical multilocus genotypes. These two strains were isolated from the same host, one from the oral cavity

and the other from a fingernail. However, it should be pointed out that the small sample sizes analysed here might not be representative of the global population of *C. fermentati*. Additional samples will be needed before a more robust conclusion can be drawn about the predominant mode of reproduction in natural populations of *C. fermentati*.

Geographical pattern of molecular variation in *C. guilliermondii*

Because of the small sample size for *C. fermentati*, our analysis of the geographical pattern of molecular variation was restricted to *C. guilliermondii*. As indicated above and shown in Fig. 2, three clones were found among the 32 strains of *C. guilliermondii*. All three clones were restricted to a specific geographical region: one was found only in the Philippines, one only in Ontario, Canada, and the third only in China. These results suggest a geographical pattern for clonal distribution in *C. guilliermondii*. Interestingly, these three clones were genetically very similar to each other, with only 1–2 nucleotide difference(s) (out of a total 3208 nucleotides) separating them. In addition, five other genotypes with one strain each (four from China and one from Toronto, Canada) differed at only one nucleotide position from the large clone from China. The small number of nucleotide differences suggests a recent common origin for these 30 strains. However, when all 32 strains of *C. guilliermondii* were analysed together using the T-PTP test, there was no statistically robust geographical pattern ($P<0.05$). The lack of a strict geographical pattern is evident in the combined phylogenetic tree (Fig. 2). Specifically, strains from none of the three regions (Ontario, the Philippines and China) were exclusively clustered together with each clade containing only strains from one specific region but not from other regions. Taken together, the evidence from the above analyses suggests the existence of dispersion among geographical populations of *C. guilliermondii*.

Are there additional divergent lineages in *C. guilliermondii*?

While strains PH144 and PH145 from the Philippines had *ITS* sequences indistinguishable from the majority of strains from China, their sequences in the *ACT1*, *EF1*, *RIBO* and *TOPOII* genes were consistently different from the other 30 *C. guilliermondii* strains (Table 1, Figs 1 and 2). The combined analysis of all five genes revealed that the mean K2P distance of the two strains from the other 30 strains of *C. guilliermondii* ($K2P$ distance 0.01274 ± 0.00035 ; mean $\pm$ SD, $n=60$ pairwise comparisons) was about 30 times the divergence among strains within the main *C. guilliermondii* clade (0.00042 ± 0.00031 , $n=435$ pairwise comparisons). These results suggest that the current species *C. guilliermondii* may contain additional evolutionarily divergent lineages.

DISCUSSION

This is believed to be the first multiple gene genealogical analysis and comparison of natural strains of *C. guilliermondii*

pattern of molecular variation, we identified several region-specific clones of *C. guilliermondii*. Samples of both species exhibited evidence of clonal reproduction and limited or no evidence for recombination. Interestingly, our results also suggested that there might be additional divergent lineages within *C. guilliermondii*.

Fig. 1. Maximum-parsimony trees for 37 strains of *C. guilliermondii* and *C. fermentati* from each of five gene fragments sequenced. (a) *ACT1*, tree length=14, consistency index (CI)=1.0; (b) *EF1*, tree length=20, CI=1.0; (c) *ITS*, tree length=7, CI=1.0; (d) *RIBO*, tree length=96, CI=0.938; (e) *TOPOII*, tree length=63, CI=0.984. The branch lengths are proportional to the amount of sequence divergence within each gene tree. The number of nucleotide changes for each branch is presented. Note the differences among genes in the scale bars: 0.5 change for (a) and (b), 0.1 change for (c), 1 change for (d) and (e).

The observed significant molecular divergence between *C. guilliermondii* and *C. fermentati* in all five genes analysed suggested that potentially useful molecular markers could be developed to distinguish strains of the two species. Such a marker system could be highly valuable in clinical microbiology laboratories. Based on the consensus DNA sequences for each of the two species, a list of diagnostic restriction polymorphic sites distinguishing *C. guilliermondii* and *C. fermentati* was generated (Table 5).

These polymorphic sites can be assayed through simple PCR reactions using primers listed in Table 2, followed by restriction enzyme digests of the PCR products and gel electrophoresis of the digested DNA. Such assays are unambiguous, rapid, economical, and ideal for clinical application.

Because *C. guilliermondii* is capable of sexual reproduction in the laboratory, we hypothesized that there should be

Table 3. Mean pairwise Kimura two-parameter distances ($\times 10^{-2}$) within and between *C. guilliermondii* and *C. fermentati*

Gene	Within <i>C. guilliermondii</i> mean $\pm$ SD (32 strains, 496 pairwise distances)	Within <i>C. fermentati</i> mean $\pm$ SD (5 strains, 10 pairwise distances)	Between <i>C. guilliermondii</i> and <i>C. fermentati</i> mean $\pm$ SD (160 pairwise distances)
<i>ACT1</i>	0.017038 $\pm$ 0.045924	0.140875 $\pm$ 0.11506	1.782904 $\pm$ 0.121011
<i>EF1</i>	0.191716 $\pm$ 0.515417	0.069204 $\pm$ 0.089342	2.76743 $\pm$ 0.150959
<i>ITS</i>	0.105557 $\pm$ 0.09864	0	0.750683 $\pm$ 0.083583
<i>RIBO</i>	0.530473 $\pm$ 1.144789	0.271602 $\pm$ 0.143100	14.132157 $\pm$ 0.583237
<i>TOPOII</i>	0.166359 $\pm$ 0.421979	0.515908 $\pm$ 0.389524	7.384368 $\pm$ 0.31236
All genes	0.192233 $\pm$ 0.403768	0.207279 $\pm$ 0.1070	5.109388 $\pm$ 0.140201

Fig. 2. One of 28 maximum-parsimony trees from the combined DNA sequences of the *ACT1*, *EF1*, *ITS*, *RIBO* and *TOPOII* genes. Tree length=204, CI=0.946; scale bar, 5 changes.

evidence for genealogical incongruence and recombination in natural populations of *C. guilliermondii*. Contrary to our expectation, we found no evidence for recombination in either *C. guilliermondii* or *C. fermentati* populations analysed here. These results contrast with those from the common yeast *C. albicans*, for which evidence for recombination in natural populations is often found despite our failure to identify its complete sexual reproductive cycle in the laboratory (Pujol *et al.*, 2005; Xu *et al.*, 1999c). It should be pointed out that the lack of evidence for recombination in the samples of *C. guilliermondii* and *C. fermentati* was not due to small sample sizes per se. In our analyses of clonality and recombination, three tests with two opposite null models, strict clonality (implemented in the incompatibility

ratio test and the PH test) and random mating (implemented in the linkage equilibrium or I_A test) were conducted and compared. In population genetic analyses, small sample size typically results in the lack of power to reject the null hypothesis. The fact that the null hypothesis of random mating in the index of association test was rejected in samples of both species suggests unambiguous evidence for clonality in the analysed samples. However, it should be emphasized that although the analysed samples are consistent with clonal reproduction in these two species, for several reasons, our data cannot exclude the possibility of recombination in either species.

First, although our samples were collected from diverse geographical regions without any *a priori* set of criteria, the samples analysed here were relatively small and may not be representative of the global populations of the two species. It is possible that samples from other geographical areas or natural environments may exhibit evidence of recombination. Analysis of additional samples is needed to identify the potential existence of such geographical or ecological-niche-specific population structures in these two species.

Second, the overall level of molecular variation is relatively limited within each of the two species. As a result, the two tests that used strict clonality as the null hypotheses may have lacked the power to reject the null hypothesis. Additional analyses using more polymorphic markers such as microsatellite loci may offer greater power to reject the hypothesis of strict clonality (Xu, 2005). Such markers should also allow a more critical examination of potentially divergent lineages within *C. guilliermondii*. However, it should be emphasized that, despite the low level of nucleotide variation within the two samples analysed here, the null hypothesis of random mating was rejected for both in the test of linkage equilibrium.

Third, the five marker genes may be functionally linked and thus significantly associated. Physically, these five genes are located on four different chromosomes (or supercontigs; Table 2) as shown in the recent assembly of a 10 × coverage genome sequencing of a *C. guilliermondii* strain by the Broad Institute http://www.broad.mit.edu/annotation/fungi/candida_guilliermondii/background.html). Only two genes, *ITS* and *TOPOII*, are located on the same chromosome (supercontig 5). This supercontig is about 1.2 million base pairs long and *ITS* and *TOPOII* are located about 860 kb

Table 4. Analyses of association among haplotypes in populations of *C. guilliermondii* and *C. fermentati*

Sample type	Index of association		Phylogenetic incompatibility	
	I_A	P	% Incompatible	P
<i>C. guilliermondii</i> (n=32)	1.238	<0.001	0.000	>0.999
<i>C. fermentati</i> (n=5)	0.831	=0.045	0.000	>0.851

Table 5. Examples of restriction site polymorphisms in fragments of five genes that distinguish *C. guilliermondii* and *C. fermentati*

For each gene, the restriction site polymorphisms were derived from consensus sequences of 32 strains of *C. guilliermondii* and 5 strains of *C. fermentati*. Only examples of restriction enzymes that cut only once or not at all are given here for each gene.

Gene	Uncut PCR fragment size (bp)	Restriction enzyme	Expected fragment size(s) of strains of <i>C. guilliermondii</i> (bp)	Expected fragment size(s) of strains of <i>C. fermentati</i> (bp)
<i>ACT1</i>	711	<i>BceAI</i>	711	299/412
		<i>BpuEI</i>	119/592	711
<i>EF1</i>	579	<i>AlwI</i>	421/158	579
		<i>KpnI</i>	517/62	579
		<i>HaeIII</i>	579	119/460
<i>ITS</i>	585	<i>Hpy188I</i>	403/182	585
		<i>BsrI</i>	585	444/141
		<i>MmeI</i>	585	503/82
<i>RIBO</i>	597	<i>AlwI</i>	494/103	597
		<i>HincII</i>	327/270	597
		<i>HgaI</i>	597	445/152
<i>TOPOII</i>	703	<i>AgeI</i>	190/513	703
		<i>BanII</i>	540/263	703
		<i>BseRI</i>	703	254/549

away from each other, almost at opposite ends of the chromosome. At present, the physical locations of the five genes are unknown in the *C. fermentati* genome. However, although physically these genes are far apart from each other in the *C. guilliermondii* genome, it is possible that they are functionally linked and interact with each other epistatically to influence fitness of strains. Epistatic interactions could allow the build-up of linkage disequilibrium in natural populations (Xu, 2004b). Analysing different marker genes should allow a direct test of this possibility.

Lastly, because of the limited genetic variation within each of the two samples, it is highly likely that even if sexual reproduction were common in populations of *C. guilliermondii* and *C. fermentati*, the analyses conducted here might not have been able to discriminate it from clonal reproduction. For example, two forms of sexual reproduction, crossing between close relatives (i.e. inbreeding) and self-fertilization, can both generate progeny with identical genotypes that would be recognized as belonging to the same asexual clones (Xu, 2005; Xu *et al.*, 2005).

We found geographically specific clones of *C. guilliermondii* in three locations (China, the Philippines, and Ontario, Canada). This result suggests that there might be large locally adapted clones in this species. The Ontario clone contained strains from the metropolitan Toronto area as well as from a small native community about 70 km west of Toronto, two economically and ecologically very different areas. The clone from China contained strains from three locations separated by hundreds to thousands of kilometres: a village in southeast China (Shitou Village, Jiangxi Province), a large city in central China (Xi'an, Shaanxi Province), and another large city in southwest China

(Chengdu, Sichuan Province). Furthermore, the close similarity in the DNA sequences among the three clones as well as the other five strains (TOR13, ZG57, ZG70, ZG248, ZG379; Fig. 2) suggests a recent common origin of the 30 *C. guilliermondii* strains (except strains PH144 and PH145).

Like most micro-organisms that show wide geographical distribution of certain genotypes but where the mechanisms for such patterns remain unknown (e.g. Kidd *et al.*, 2005; Xu, 2004a; Xu *et al.*, 1999b, 2005), the reasons for the wide distribution of some of the genotypes in *C. guilliermondii* are also unknown at present. However, there are several possibilities. First, human travelling in the past several decades might have contributed to the spread of certain genotypes. Second, the populations of *C. guilliermondii* from Asia and North America might have been the result of a severe population bottleneck associated with prehistoric humans when they migrated from Africa to other parts of the world a few hundred thousand years ago. Third, unlike *C. albicans*, for which mammals are the primary host, *C. guilliermondii* can be found in diverse environments such as water and soil, as well as a variety of animals (Odds, 1988; Barnett *et al.*, 2000). Therefore, there could be dispersal through these environmental sources. All three possibilities could have contributed to clonal dispersal in *C. guilliermondii*.

The identification of additional divergent lineages within the current species of *C. guilliermondii* further supports the notion that clinical strains previously identified as *C. guilliermondii* consist of a group of closely related phylogenetic species. At present, the medical significance of the species and divergent clones is unknown. However, using

the information obtained here, and with the availability of a genome sequence (http://www.broad.mit.edu/annotation/fungi/candida_guilliermondii/background.html), the analysis of additional clinical and environmental strains of *C. guilliermondii* using these and additional genetic markers should establish a solid evolutionary framework to allow a comprehensive examination of the epidemiology, pathogenesis, and control and treatment of *C. guilliermondii* infections.

ACKNOWLEDGEMENTS

We thank many people who have contributed to the strain collections: Dr Glen Bulmer for the six strains from the Philippines, Laura Montour for the three strains from the First Nations Reserve in Ontario, Dr Debby Yamamura for the 12 strains from Toronto, Ontario, and Dr Wiley Schell of Duke University for the two strains from Tanzania. We thank Heather Yoell for comments. This work was funded by the Premier's Research Excellence Award, Genome Canada, and the Natural Science and Engineering Research Council (NSERC) of Canada.

REFERENCES

- Agapow, P. M. & Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Mol Ecol Notes* 1, 101–102.
- Bai, F. Y., Liang, H. Y. & Jia, J. H. (2000). Taxonomic relationships among the taxa in the *Candida guilliermondii* complex, as revealed by comparative electrophoretic karyotyping. *Int J Syst Evol Microbiol* 50, 417–422.
- Barnett, J. A., Payne, R. W. & Yarrow, D. (2000). *Yeast Characteristics and Identification*, 3rd edn. Cambridge: Cambridge University Press.
- Doi, M., Homma, M., Chindamporn, A. & Tanaka, K. (1992). Estimation of chromosome number and size by pulsed-field gel electrophoresis (PFGE) in medically important *Candida* species. *J Gen Microbiol* 138, 2243–2251.
- Dos Santos, A. L. & Soares, R. M. (2005). *Candida guilliermondii* isolated from HIV-infected human secretes a 50 kDa serine proteinase that cleaves a broad spectrum of proteinaceous substrates. *FEMS Immunol Med Microbiol* 43, 13–20.
- Faith, D. L. (1991). Cladistic permutation tests for monophyly and nonmonophyly. *Syst Zool* 40, 366–375.
- Farris, J. S., Kallersjo, M., Kluge, A. G. & Bult, C. (1995). Testing significance of incongruence. *Cladistics* 10, 315–319.
- Imshenetskii, A. A., Solntseva, L. I. & Afanas'eva, L. N. (1979). Biomass accumulation by polyploid forms of *Candida guilliermondii* yeasts. *Mikrobiologiya* 48, 418–422.
- Kam, A. & Xu, J. (2002). Diversity of commensal yeasts from within and between healthy hosts. *Diag Microbiol Infect Dis* 43, 19–28.
- Kidd, S. E., Guo, H., Bartlett, K., Xu, J. & Kronstad, J. W. (2005). Comparative gene genealogies indicate that two clonal lineages of *Cryptococcus gattii* in British Columbia resemble strains from other geographical areas. *Eukaryot Cell* 4, 1629–1638.
- Krcmery, V. & Barnes, A. J. (2002). Non-*Candida albicans* spp. causing fungaemia: pathogenicity and antifungal resistance. *J Hosp Infect* 50, 243–260.
- Maiden, M. C., Bygraves, J. A., Feil, E. & 10 other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 95, 3140–3145.
- McEwen, J. G., Taylor, J. W., Carter, D. & 8 other authors (2000). Molecular typing of pathogenic fungi. *Med Mycol* 38, 189–197.
- Odds, F. C. (1988). *Candida and Candidiasis*. London: Baillière Tindall.
- Pujol, C., Dodgson, A. & Soll, D. R. (2005). Population genetics of ascomycetes pathogenic to humans and animals. In *Evolutionary Genetics of Fungi*, pp. 149–188. Edited by J. Xu. Norwich, UK: Horizon Scientific Press.
- Sandven, P. (2000). Epidemiology of candidemia. *Rev Iberoam Micol* 17, 73–81.
- San Millan, R. M., Wu, L. C., Salkin, I. F. & Lehmann, P. F. (1997). Clinical isolates of *Candida guilliermondii* include *Candida fermentati*. *Int J Syst Bacteriol* 47, 385–393.
- Sullivan, D. J., Westerneng, T. J., Haynes, K. A., Bennett, D. E. & Coleman, D. C. (1995). *Candida dubliniensis* sp. nov.: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals. *Microbiology* 141, 1507–1521.
- Swofford, D. L. (2004). PAUP* – Phylogenetic Analysis Using Parsimony* and Other Methods, version 4. Sunderland, MA: Sinauer Associates.
- Tavanti, A., Davidson, A. D., Gow, N. A., Maiden, M. C. & Odds, F. C. (2005). *Candida orthopsilosis* and *Candida metapsilosis* spp. nov. to replace *Candida parapsilosis* groups II and III. *J Clin Microbiol* 43, 284–292.
- Tey, R., Yan, Z., Han, S., Li, X., Lazazzera, K., Sun, S. & Xu, J. (2003). Ecology and epidemiology of drug resistance in human pathogenic yeasts. *Res Adv Microbiol* 3, 67–85.
- Warren, N. G. & Hazen, K. C. (1999). *Candida*, *Cryptococcus*, and other yeasts of medical significance. In *Manual of Clinical Microbiology*, 7th edn, pp. 1184–1199. Edited by P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover & R. H. Tenover. Washington, DC: American Society for Microbiology.
- Xu, J. (2004a). Gene genealogical analyses of human fungal pathogens. In *Pathogenic Fungi: Structural Biology and Taxonomy*, pp. 299–318. Edited by G. San-Blas & R. Calderone. Norwich: Horizon Scientific Press.
- Xu, J. (2004b). The prevalence and evolution of sex in microorganisms. *Genome* 47, 775–780.
- Xu, J. (2005). Fundamentals of fungal molecular population genetic analyses. *Evolutionary Genetics of Fungi*, pp. 87–116. Edited by J. Xu. Norwich: Horizon Scientific Press.
- Xu, J. (2006). Microbial ecology in the age of genomics and metagenomics: concepts, tools and recent advances. *Mol Ecol* (in press).
- Xu, J. & Mitchell, T. G. (2003a). Geographic differences in human oral yeast flora. *Clin Infect Dis* 36, 221–224.
- Xu, J. & Mitchell, T. G. (2003b). Comparative gene genealogical analyses of strains of serotype AD identify recombination in populations of serotypes A and D in the human pathogenic yeast *Cryptococcus neoformans*. *Microbiology* 149, 2147–2154.
- Xu, J., Boyd, C. M., Livingstone, E., Meyer, W., Madden, J. F. & Mitchell, T. G. (1999a). Species and genotypic diversities and similarities of pathogenic yeasts colonizing women. *J Clin Microbiol* 37, 3835–3843.
- Xu, J., Vilgalys, R. & Mitchell, T. G. (1999b). Lack of genetic differentiation between two geographic samples of *Candida albicans* isolated from patients infected with the human immunodeficiency virus. *J Bacteriol* 181, 1369–1373.
- Xu, J., Mitchell, T. G. & Vilgalys, R. (1999c). PCR-RFLP analyses reveal both extensive clonality and local genetic differentiation in *Candida albicans*. *Mol Ecol* 8, 59–74.

Xu, J., Ramos, A. R., Vilgalys, R. & Mitchell, T. G. (2000a). Clonal and spontaneous origins of fluconazole resistance in *Candida albicans*. *J Clin Microbiol* **38**, 1214–1220.

Xu, J., Vilgalys, R. & Mitchell, T. G. (2000b). Multiple gene genealogies reveal recent dispersion and hybridization in the human pathogenic fungus *Cryptococcus neoformans*. *Mol Ecol* **9**, 1471–1481.

Xu, J., Luo, G., Vilgalys, R. J., Brandt, M. E. & Mitchell, T. G. (2002). Multiple origins of hybrid strains of *Cryptococcus neoformans* with serotype AD. *Microbiology* **148**, 203–212.

Xu, J., Cheng, M., Tan, Q. & Pan, Y. (2005). Molecular population genetics of basidiomycete fungi. In *Evolutionary Genetics of Fungi*, pp. 221–252. Edited by J. Xu. Norwich, UK: Horizon Scientific Press.