The yeast *Candida albicans* has a clonal mode of reproduction in a population of infected human immunodeficiency virus-positive patients

(genetic polymorphism/clonal theory/AIDS/population structure)

Claude Pujol*, Jacques Reynes†, François Renaud‡, Michel Raymondu, Michel Tibayrenc†, Francisco J. Ayala†, François Janbon†, Michèle Malliet‡, and Jean-Marie Bastide*,**

*Laboratoire d’Immunologie et Parasitologie, Faculté de Pharmacie, 15, Avenue Charles Flahault, 34060 Montpellier Cédex 1, France; †Maladies Infectieuses A, Hôpital Gui de Chauliac, Centre Hospitalier Universitaire, 34059 Montpellier Cédex, France; ‡Laboratoire de Parasitologie Comparée (Unité de Recherche Associate 690) and †Laboratoire Génétique et Environnement, Institut des Sciences de l’Évolution (Unité de Recherche Associated 1557, Centre National de la Recherche Scientifique, Université des Sciences et Techniques du Languedoc, Place Eugène Bataillon, 34005 Montpellier Cédex 5, France; ‡Génétique Moléculaire des Parasites et des Vecteurs (Unité Mixte de Recherche 9326, Centre National de la Recherche Scientifique-Institut Français de Recherche Scientifique pour le Développement en Coopération), Institut Français de Recherche Scientifique pour le Développement en Coopération, BP 5045, 34032 Montpellier Cédex 1, France; and †Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92617

Contributed by Francisco J. Ayala, July 19, 1993

ABSTRACT To ascertain the population structure of *Candida albicans*, we have carried out a multilocus enzyme electrophoresis study based on the analysis of 21 gene loci. We have thus characterized 55 strains isolated one each from 55 human immunodeficiency virus-positive patients. There is considerable polymorphism among the strains. A population-genetic analysis indicates that the two fundamental consequences of sexual reproduction (i.e., segregation and recombination) are apparently absent in this population of *C. albicans*. The population structure of *C. albicans* appears to be clonal, a state of affairs that has important medical and biological consequences.

*Candida albicans* (Robin) Berkhout (1923) mycoses are responsible for severe mucosal or systemic infections in immunocompromised patients. In patients with AIDS, oropharyngeal candidiasis is the most common fungal infection and is predictive of the development of AIDS (1, 2); antifungal treatments are often initially successful but relapses are common (3). The affinity of *C. albicans* mycoses for human immunodeficiency virus-positive (HIV+) patients has led to the search for improved epidemiological knowledge of the nature of this opportunistic pathogen is investigated following the population genetic tests proposed by Tibayrenc et al. (12, 13).

Candida albicans (Robin) Berkhout (1923) mycoses are responsible for severe mucosal or systemic infections in immunocompromised patients. In patients with AIDS, oropharyngeal candidiasis is the most common fungal infection and is predictive of the development of AIDS (1, 2); antifungal treatments are often initially successful but relapses are common (3). The affinity of *C. albicans* mycoses for human immunodeficiency virus-positive (HIV+) patients has led to the search for improved epidemiological knowledge of the nature of this opportunistic pathogen is investigated following the population genetic tests proposed by Tibayrenc et al. (12, 13).

**Materials and methods**

*C. albicans* strains were isolated, one per person, from 55 HIV+ patients suffering from oropharyngeal candidiasis, according to ref. 3. The samples were taken between January 1990 and June 1991 in an AIDS unit (Clinique des Maladies Infectieuses A, Montpellier, France). Starch gel electrophoresis and enzymatic assays were performed following described protocols (15). Data were obtained for 19 enzymatic activities: malate dehydrogenase (EC 1.1.1.37), glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49), sorbitol dehydrogenase (EC 1.1.1.14), isocitrate dehydrogenase (EC 1.1.1.42), alcohol dehydrogenase (EC 1.1.1.1), superoxide dismutase (EC 1.15.1.1), hexokinase (EC 2.7.1.1), pyruvate kinase (EC 2.7.1.40), aspartate aminotransferase (EC 2.6.1.1), phosphoglucomutase (EC 5.4.2.2), esterase (EC 3.1.1.1), leucine aminopeptidase (EC 3.4.11.1), and acid phosphatase (EC 3.1.1.1). Leucine aminopeptidase (EC 3.4.11.1), peptidase 1 (EC 3.4.13.18; substrate, Val-Leu), peptidase 2 (EC 3.4.11.4; substrate, Leu-Gly-Gly), peptidase 3 (EC 3.4.13.9; substrate, Phe-Pro), aldolase (EC 4.1.2.13), fumarase (EC 4.2.1.2), mannose-6-phosphate isomerase (EC 5.3.1.18), glucose-6-phosphate isomerase (EC 5.3.1.9). Malate dehydrogenase and hexoki-
vase enzymatic activities were each expressed by 2 loci: Mdh-1 and Mdh-2, Hk-1 and Hk-2. Thus, data were obtained for 21 genetic loci.

RESULTS

Of the 21 loci investigated, 13 (62%) exhibit variability (Table 1; Fig. 1), yielding 41 different enzyme profiles. Of these multilocus patterns, 39 are represented by one strain each, 1 is represented by 3 strains, and 1 is represented by 13 strains. The average number of alleles among the variable loci is 2.85, with a range from 2 to 5. The observed heterozygosity is 0.168 ± 0.018, with a range from 0.048 to 0.333 (Table 1). We have measured genetic divergence between the clones by Nei’s genetic distance (17), which estimates the number of codon differences per locus between two clones. The mean value for the 820 possible pairwise comparisons is $D = 0.168 ± 0.006$, with a range from 0.013 to 0.430.

The segregation tests depend on an allelic interpretation of the isozyme patterns. The results of a test for ascertaining whether the observed genotypic frequencies are consistent with those expected from random mating are shown in Table 1. The frequencies of single-locus genotypes are grossly inconsistent with the expected Hardy-Weinberg frequencies at several loci as well as for all loci combined. Some genotypes that would be expected at high frequencies if the population were panmictic do not appear at all in the sample; for example, the genotype 1/2 at the Mdh-1 locus (expected number, 10.68; $\chi^2 = 13.25; P < 0.001$). If the population were panmictic, the expected heterozygosity for all loci considered together would be 0.204, whereas the observed heterozygosity is 0.168 ± 0.018.

Recombination tests are independent of the assumption of ploidy level and, hence, more robust in this respect than segregation tests. Each distinguishable enzyme pattern at a given locus is considered a distinct genotype (without pre-judging its allelic makeup). The expected frequency of a given multilocus genotype is simply the product of the observed frequencies of the corresponding single-locus genotypes. Strong indications of linkage disequilibrium are evident in the present sample (Table 2). For example, the probability of observing the most common multilocus genotype as many times as actually observed (observed size, 13; expected size, 0.12) is $3 \times 10^{-22}$. It is apparent that genetic recombination is far from random in this sample.

DISCUSSION

The commonly accepted notion that C. albicans is diploid is supported by the observation of typical heterozygous patterns in the enzyme assays (Fig. 1). This result agrees with the results presented by Caugant and Sandven (10) showing

| Table 1. Genetic variability and nonrandom segregation in 55 strains of C. albicans |
|-----------------|----------|-----------|-----------|-----------|---------------|
| Loci           | Genotypes (frequencies) | Hardy–Weinberg test, level of significance | Over-represented multilocus genotypes |
|                |                      | P1       | P2       |             |               |
| Acn            | 1/1 (0.96), 2/2 (0.04) | 0.001    | 0.001    | 1/1        | 1/1           |
| GSPd           | 1/1 (0.55), 2/2 (0.07), 1/2 (0.38) | NS       | NS       | 1/1        | 1/2           |
| Gpi            | 1/1 (0.98), 1/2 (0.02) | NS       | NS       | 1/1        | 1/1           |
| Hk-1           | 1/1 (0.45), 2/2 (0.18), 1/2 (0.24), 2/3 (0.13) | 0.001    | NS       | 1/1        | 1/2           |
| Hk-2           | 1/1 (0.49), 2/2 (0.08), 1/2 (0.34), 1/3 (0.15) | NS       | NS       | 1/2        | 1/5           |
| Lap            | 1/1 (0.46), 2/2 (0.05), 1/2 (0.47), 1/2 (0.02) | NS       | NS       | 1/2        | 1/1           |
| Mdh-1          | 1/1 (0.89), 2/2 (0.11) | 0.001    | 0.001    | 1/1        | 1/1           |
| Mdh-2          | 1/1 (0.84), 2/2 (0.02), 1/2 (0.14) | NS       | NS       | 1/1        | 1/1           |
| Mpi            | 1/1 (0.09), 2/2 (0.07), 2/3 (0.04), 1/2 (0.42), 1/3 (0.11), 1/4 (0.14), 1/5 (0.07), 2/3 (0.04), 2/5 (0.02) | 0.001    | NS       | 1/2        | 1/2           |
| Pep-1          | 1/1 (0.98), 2/2 (0.02), 1/2 (0.02) | NS       | NS       | 1/1        | 1/1           |
| Pep-2          | 1/1 (0.53), 2/2 (0.04), 3/2 (0.02), 4/2 (0.02), 1/2 (0.04), 1/3 (0.27), 1/5 (0.04), 2/3 (0.02), 3/4 (0.62) | NS       | NS       | 1/1        | 1/1           |
| Pep-5          | 1/1 (0.53), 2/2 (0.18), 3/2 (0.05), 4/2 (0.02), 1/2 (0.09), 1/3 (0.09), 2/3 (0.04) | 0.001    | 0.001    | 1/1        | 1/3           |
| Pk             | 1/1 (0.65), 2/2 (0.13), 1/3 (0.22) | 0.005    | NS       | 1/1        | 1/1           |
| Fooled data    |                      | <0.005   | <0.005   | <0.005     | <0.005       |

For each locus, allele $I$ codes for the fastest migrating electromorph. Genotypes (and their frequencies, in parentheses) are shown at each locus. The level of significance is given for each Hardy–Weinberg test; NS, not significant. Fisher’s exact test is used for the individual loci, after combining the rarer alleles into one class when more than two alleles are observed. Tests for pooled data follow Fisher’s method for combining independent test results (16). P1 is for the whole data set; the two predominant genotypes are removed from the P2 tests. The two over-represented multilocus genotypes are each observed in more than one strain; the left one in 13 strains, and the right one in 3 strains. Loci encode the following enzymes: Acn, aspartate aminotransferase; GSPd, glucose-6-phosphate 1-dehydrogenase; Gpi, glucose-6-phosphate isomerase; Hk, hexokinase; Lap, leucine aminopeptidase; Mdh, malate dehydrogenase; Mpi, mannose-6-phosphate isomerase; Pep, peptidase; Pk, pyruvate kinase.
typical heterozygous patterns for four enzyme loci in C. albicans. "Heterozygous" band patterns may also arise owing to either aneuploidy or gene duplication. These possibilities are unlikely in our case on the grounds that (i) duplications would need to be postulated at many loci (11 of 21 loci in our sample of 55 individuals) and (ii) no triplications have been observed as would be expected on occasion if aneuploidy were prevalent. Diploidy is the most parsimonious explanation for our observations.

Our results support the hypothesis that C. albicans has, at least in the population investigated in this work, a predominately clonal mode of propagation. Other potential factors that may generate deviations from the patterns expected in a panmictic population are geographical subdivision, happenstance linkage of the 13 polymorphic loci on the same chromosome, natural selection, and self-fertilization.

Geographical isolation may be associated with different allelic frequencies in different populations, even if each separate population is panmictic. When samples from different local populations are combined, this may result in apparent departures from Hardy–Weinberg expectations (particularly a deficit of heterozygotes, the Wahlund effect) and linkage disequilibrium. Geographical subdivision is unlikely to account for our results. First, the present sample was collected in a restricted geographical area (Montpellier, France), which reduces this possible source of error. At the same geographical scale and with comparable methods of analysis, human populations do not exhibit departures from panmictia (12). Second, although several loci show a deficit of heterozygotes, this is not always so—Mpi, for example, exhibits an excess of heterozygotes.

Linkage disequilibrium may be present in random mating populations if the loci are physically proximal on the same chromosome, whenever populations with different allelic frequencies have combined in the recent past, or when natural selection favors particular multilocus genotypic arrays and rejects others. The loci in our study were chosen without knowledge of their chromosomal location. There are eight pairs of chromosomes in C. albicans (19). The probability that all 13 polymorphic loci are located on the same chromosome is small ($P < 10^{-10}$); most likely they are scattered over several chromosomes. Even if some loci are on the same chromosome, crossing over makes it unlikely that multilocus linkage disequilibrium will be sustained in natural populations, unless the loci are all located within a small DNA segment (20).

If natural selection favors certain allelic combinations, or certain multilocus associations, it would generate departures from both Hardy–Weinberg expectations and linkage disequilibrium. Assume, for example, that two common multilocus genotypes (the two highly overrepresented in our sample) are strongly favored by natural selection within a sexual population. The loci positively selected may not necessarily be the ones evidences by our enzyme assays but may be others tightly linked to them ("hitchhiking"). In any case, natural selection can sustain linkage disequilibrium only by eliminating the genotypes arisen by recombination. As the number of loci in disequilibrium increases above a few, the fraction of the population that needs to be eliminated (the "genetic load" of the population) soon becomes unbearable (12). Furthermore, linkage disequilibrium persists in our data even when the two overrepresented genotypes are removed from the analysis (Table 2).

Self-fertilization results in deficiency of heterozygotes and linkage disequilibrium. It might account for some of our results. But the presence of heterozygotes at several loci in the two overrepresented genotypes and the excess heterozygosity observed at the Mpi locus suggest that departures from panmixia are not in our case conveniently explained by self-fertilization.

One might infer that the two predominant genotypes observed in the sample are the only ones generated by clonality, while the others (each observed only once) would be the result of recombinations. But, as pointed out, considerable linkage disequilibrium persists even when the two predominant genotypes are removed from the tests (Table 2).

We conclude that clonal reproduction is the most parsimonious overall explanation for the results. A clonal population structure does not imply that recombination is totally absent, but only that it is too rare an event to break a prevalent pattern of clonality (13). Trypanosoma brucei can undergo mating in the laboratory (21), and this might also occur in the wild. Yet T. brucei natural populations have clonal structures (12, 13), which suggests that recombination plays a role, if at all, only on an evolutionary scale. Escherichia coli, as well as other species of bacteria, exhibit clonal population structures, even though occasional recombination occurs (22, 23). Similarly, the present data do not rule out some genetic recombination in C. albicans but show that recombination is severely restricted, at least in the population we have surveyed.

Tibayrenc et al. (12) did not find any clear indication of clonal propagation in C. albicans. A similar conclusion has recently been reached by Caugant and Sandven (10). It will be important to investigate other populations of C. albicans, derived from various epidemiological foci, in order to ascertain whether or not the clonal hypothesis is generalizable to the whole taxon C. albicans.

The genetic distances observed for C. albicans are lower than those for Trypanosoma cruzi (11) ($D = 0.757 \pm 0.478$; range, 0.107–0.215). Nevertheless, the genetic divergence between clones is substantial—comparable to that existing between different mammal species of the same genus, for example. Genetic differentiation of such magnitude is likely to be reflected in epidemiological and pathological differences between clones. This, in turn would seem to call for separate investigation of the medical attributes of clones.

One possibility to keep in mind is that some clonal lineages may have widespread distribution, as is the case in some parasitic protozoa (11–13) and bacteria (22, 23). Genetically differentiated clones, particularly any ones that might have worldwide distribution, should be separately investigated for relevant medical features such as pathogenicity, virulence, or resistance to antiparasitic agents. A definite association between the genetic makeup of clones and their virulence has been recently uncovered in Toxoplasma gondii (24), an

Table 2. Linkage disequilibrium in *C. albicans*

<table>
<thead>
<tr>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$e$</th>
<th>$f$</th>
<th>$\text{Var}(d_{ij})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.03</td>
<td>0.09</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>0.02</td>
<td>0.05</td>
<td>0.03</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Levels of significance for nonrandom association between loci (linkage disequilibrium) under the null hypothesis of random recombination. $d_1$ = combinatorial probability of sampling a given genotype as often or more than actually observed ($d_2$ = two values refer to the two genotypes observed, respectively, in 13 and 3 strains); $d_2$ = probability of observing any given genotype as often as or more than the most common genotype actually observed; $e$ = probability of observing as few or fewer genotypes than actually observed; $f$ = probability that linkage disequilibrium is as high as observed; and $\text{Var}(d_{ij})$ have been previously used for evidencing clonality in parasitic protozoa (12, 13). $\text{Var}(d_{ij})$ is an additional test for nonrandom association between the $n(n-1)/2$ pairs of individuals sampled (18); significance is based on the same Monte Carlo simulations used for $d_2$. $P_1$ = total sample of 55 strains; $P_2$ = sample of 39 strains obtained by removing the two predominant genotypes shown in Table 1; only $f$ and $\text{Var}(d_{ij})$ can be performed, since removal of the only two repeated genotypes in the sample makes tests $d_1$, $d_2$, and $e$ meaningless.

If natural selection favors certain allelic combinations, or certain multilocus associations, it would generate departures from both Hardy–Weinberg expectations and linkage disequilibrium. Assume, for example, that two common multilocus genotypes (the two highly overrepresented in our sample) are strongly favored by natural selection within a sexual population. The loci positively selected may not necessarily be the ones evidenced by our enzyme assays but may be others tightly linked to them ("hitchhiking"). In any case, natural selection can sustain linkage disequilibrium only by eliminating the genotypes arisen by recombination. As the number of loci in disequilibrium increases above a few, the fraction of the population that needs to be eliminated (the "genetic load" of the population) soon becomes unbearable (12). Furthermore, linkage disequilibrium persists in our data even when the two overrepresented genotypes are removed from the analysis (Table 2).

Self-fertilization results in deficiency of heterozygotes and linkage disequilibrium. It might account for some of our results. But the presence of heterozygotes at several loci in the two overrepresented genotypes and the excess heterozygosity observed at the Mpi locus suggest that departures from panmixia are not in our case conveniently explained by self-fertilization.

One might infer that the two predominant genotypes observed in the sample are the only ones generated by clonality, while the others (each observed only once) would be the result of recombinations. But, as pointed out, considerable linkage disequilibrium persists even when the two predominant genotypes are removed from the tests (Table 2).

We conclude that clonal reproduction is the most parsimonious overall explanation for the results. A clonal population structure does not imply that recombination is totally absent, but only that it is too rare an event to break a prevalent pattern of clonality (13). Trypanosoma brucei can undergo mating in the laboratory (21), and this might also occur in the wild. Yet T. brucei natural populations have clonal structures (12, 13), which suggests that recombination plays a role, if at all, only on an evolutionary scale. Escherichia coli, as well as other species of bacteria, exhibit clonal population structures, even though occasional recombination occurs (22, 23). Similarly, the present data do not rule out some genetic recombination in C. albicans but show that recombination is severely restricted, at least in the population we have surveyed.

Tibayrenc et al. (12) did not find any clear indication of clonal propagation in C. albicans. A similar conclusion has recently been reached by Caugant and Sandven (10). It will be important to investigate other populations of C. albicans, derived from various epidemiological foci, in order to ascertain whether or not the clonal hypothesis is generalizable to the whole taxon C. albicans.

The genetic distances observed for C. albicans are lower than those for Trypanosoma cruzi (11) ($D = 0.757 \pm 0.478$; range, 0.107–0.215). Nevertheless, the genetic divergence between clones is substantial—comparable to that existing between different mammal species of the same genus, for example. Genetic differentiation of such magnitude is likely to be reflected in epidemiological and pathological differences between clones. This, in turn would seem to call for separate investigation of the medical attributes of clones.

One possibility to keep in mind is that some clonal lineages may have widespread distribution, as is the case in some parasitic protozoa (11–13) and bacteria (22, 23). Genetically differentiated clones, particularly any ones that might have worldwide distribution, should be separately investigated for relevant medical features such as pathogenicity, virulence, or resistance to antiparasitic agents. A definite association between the genetic makeup of clones and their virulence has been recently uncovered in Toxoplasma gondii (24), an
opportunistic pathogen for which Tibayrenc et al. (13) had anticipated a clonal population structure on the basis of population genetic tests applied to earlier data available in the published literature. Association between particular genetic makeups and medical characteristics would facilitate select- ing effective methods for control of C. albicans, an opportunistic pathogen that may very well become greater of public health significance owing to its pathological expression in immunocompromised patients.