Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 9456–9459, October 1993
Evolution

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 RAYMOND§, MICHEL TIBAYRENC¶, FRANCISCO J. AYALA‖, FRANÇOIS JANBON†, MICHÈLE MALLIÉ*, 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 Associée 698) and §Laboratoire Génétique et Environnement, Institut des Sciences de l'Evolution (Unité de Recherche Associée 327), Centre National de la Recherche Scientifique, Université des Sciences et Techniques du Languedoc, Place Eugène Bataillon, 34095 Montpellier Cédex 5, France; ¶Génétique Moléculaire des Parasites et des Vecteurs (Unité Mixte de Recherche 9926, 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 92717

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 responsible strains (4).

No sexual cycle is known for *C. albicans*, and the chromosomes of this species during metaphase have never been observed. Consequently, the ploidy status of the species has not been determined by the conventional method of direct chromosome visualization (see ref. 5 for review), although it is currently accepted that *C. albicans* is diploid or close to diploid (5–10). Diploidy suggests the existence of a sexual cycle, at least in the past if it has now disappeared (11). However, the only mode of reproduction known for *C. albicans* is clonal (which is the routine procedure by which laboratories maintain and multiply this species). Consequently, it is generally accepted that *C. albicans* does not possess a sexual cycle. Nevertheless, the question of this pathogen's population structure is far from settled (10, 12). This may seem surprising, for the matter is of medical relevance; the strategies for developing curative treatments are different for clonal and sexual organisms (13).

Tibayrenc *et al.* (13) have proposed population genetic methods for investigating the prevailing mode of reproduction of an organism in nature, particularly appropriate for protozoa and other microorganisms. The two fundamental genetic consequences of sexual reproduction are segregation of alleles at a given locus and recombination of genotypes between loci. Lack of evidence for segregation and recombination in informative data is circumstantial evidence that sexual reproduction is lacking or restricted. To ascertain the prevailing mode of reproduction of human parasitic protozoa, and of several pathogenic yeasts, Tibayrenc *et al.* (12, 13) applied seven criteria to published genetic data and concluded that the majority of parasitic protozoa preferentially propagate clonally. The data published by Lehmann *et al.* (14) on the polymorphism of eight enzyme systems were, nevertheless, found insufficient to reach any definite conclusion on the population structure of *C. albicans* (12). More recently, Caugant and Sandven (10) have investigated 10 gene loci by enzyme electrophoresis without obtaining any clear evidence for clonality. These authors have emphasized the need for additional studies before rejecting clonality in *C. albicans*.

In the present work, we report the enzyme polymorphisms observed for 21 gene loci in 55 strains of *C. albicans* isolated from HIV^+ patients. The prevailing mode of reproduction in 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), 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 5.3.1.8), glucose-6-phosphate isomerase (EC 5.3.1.9). Malate dehydrogenase and hexoki-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

**To whom reprint requests should be addressed.

nase 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 $D = 0.168 \pm 0.006$, 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 prejudging 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 3×10^{-23}. It is apparent that genetic recombination is far from random in this sample.

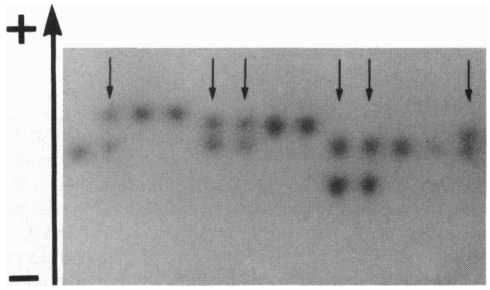

FIG. 1. Starch gel electrophoresis showing the enzyme phenotype for *Mpi* in several *C. albicans* strains. Some strains show two-banded patterns that are typical of heterozygotes for monomeric enzymes in diploid organisms.

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		
Aat	*1/1* (0.96), *2/2* (0.04)	0.001	0.001	*1/1*	*1/1*
G6Pd	*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* (0.02), *1/2* (0.34), *1/3* (0.15)	NS	NS	*1/2*	*1/3*
Lap	*1/1* (0.46), *2/2* (0.05), *1/2* (0.47), *1/3* (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), *3/3* (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*	*1/4*
Pep-1	*1/1* (0.96), *2/2* (0.02), *1/2* (0.02)	NS	NS	*1/1*	*1/1*
Pep-2	*1/1* (0.53), *2/2* (0.04), *3/3* (0.02), *4/4* (0.02), *1/2* (0.04), *1/3* (0.27), *1/5* (0.04), *2/3* (0.02), *3/4* (0.02)	NS	NS	*1/1*	*1/1*
Pep-3	*1/1* (0.53), *2/2* (0.18), *3/3* (0.05), *4/4* (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*
	Pooled data	<0.005	<0.005		

For each locus, allele *1* 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 were 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: *Aat*, aspartate aminotransferase; *G6Pd*, 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.

Table 2. Linkage disequilibrium in *C. albicans*

	d1	*d2*	*e*	*f*	Var(d_{ij})
P1	3×10^{-23}; 10^{-9}	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
P2	—	—	—	$<10^{-4}$	$<10^{-4}$

Levels of significance for nonrandom association between loci (linkage disequilibrium) under the null hypothesis of random recombination. *d1* = combinatorial probability of sampling a given genotype as often as or more than actually observed (the two values refer to the two genotypes observed, respectively, in 13 and 3 strains); *d2* = probability of observing any 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; *d2*, *e*, and *f* are based on Monte Carlo simulations with 10^4 runs. *d1*, *d2*, *e*, and *f* have been previously used for evidencing clonality in parasitic protozoa (12, 13). 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 *d2*. P1 = total sample of 55 strains; P2 = sample of 39 strains obtained by removing the two predominant genotypes shown in Table 1; only *f* and Var(d_{ij}) can be performed, since removal of the only two repeated genotypes in the sample makes tests *d1*, *d2*, and *e* meaningless.

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 predominantly 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 panmixia (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 equilibrium. 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 recombination. 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, 0.017–2.015). 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 antifungal 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 selecting effective methods for control of *C. albicans*, an opportunistic pathogen that may very well become of greater public health significance owing to its pathological expression in immunocompromised patients.

1. Klein, R. S., Harris, C. A., Small, C. B., Moll, B., Lesser, M. & Friedland, G. H. (1984) *N. Engl. J. Med.* **311,** 354–358.
2. Dodd, C. L., Greenspan, D., Katz, M. H., Westenhouse, J. L., Feigal, D. W. & Greenspan, J. S. (1991) *AIDS* **5,** 1339–1343.
3. Reynes, J., Mallié, M., Janbon, F. & Bastide, J.-M. (1992) *Pathol. Biol.* **40,** 513–517.
4. Odds, F. C., Schmid, J. & Soll, D. R. (1990) in *Mycoses in AIDS Patients*, eds. Vanden Bossche, H., Mackenzie, D., Cauwenbergh, G., Van Cutsem, J. & Drouhet, E. (Plenum, New York), pp. 67–74.
5. Scherer, S. & Magee, P. T. (1990) *Microbiol. Rev.* **54,** 226–241.
6. Poulter, R., Hanrahan, V., Jeffery, K., Markie, D., Shepherd, M. G. & Sullivan, P. A. (1982) *J. Bacteriol.* **152,** 969–975.
7. Riggsby, W. S., Torres-Bauza, L. J., Wills, J. W. & Townes, T. M. (1982) *Mol. Cell. Biol.* **2,** 853–862.
8. Whelan, W. L., Partridge, R. M. & Magee, P. T. (1980) *Mol. Gen. Genet.* **180,** 107–113.
9. Whelan, W. L. & Magee, P. T. (1981) *J. Bacteriol.* **145,** 896–903.
10. Caugant, D. A. & Sandven, P. (1993) *J. Clin. Microbiol.* **31,** 215–220.
11. Tibayrenc, M., Ward, P., Moya, A. & Ayala, F. J. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 115–119.
12. Tibayrenc, M., Kjellberg, F., Arnaud, J., Oury, B., Brenière, S. F., Dardé, M.-L. & Ayala, F. J. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 5129–5133.
13. Tibayrenc, M., Kjellberg, F. & Ayala, F. J. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2414–2418.
14. Lehmann, P. F., Kemker, B. J., Hsiao, C.-B. & Dev, S. (1989) *J. Clin. Microbiol.* **27,** 2514–2521.
15. Pujol, C., Reynes, J., Renaud, F., Mallié, M. & Bastide, J.-M. (1993) *J. Mycol. Med.* **3,** 14–19.
16. Fisher, R. A. (1985) in *The Statistics of Natural Selection on Animal Populations*, ed. Manly, B. F. J. (Chapman & Hall, London), pp. 432–433.
17. Nei, M. (1972) *Am. Nat.* **106,** 283–292.
18. Brown, A. H. D. & Feldman, M. W. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 5913–5916.
19. Wickes, B., Staudinger, J., Magee, B. B., Kwon-Chung, K.-J., Magee, P. T. & Scherer, S. (1991) *Infect. Immun.* **59,** 2480–2484.
20. Hartl, D. L. & Clark, A. G. (1989) *Principles of Population Genetics* (Sinauer, Sunderland, MA), 2nd Ed., pp. 45–57.
21. Jenni, L., Matri, S., Schweitzer, J., Betschart, B., Le Page, R. W. H., Wells, J. M., Tait, A., Paindavoine, P., Pays, H. & Steinert, M. (1986) *Nature (London)* **322,** 173–175.
22. Selander, R. K. & Levin, B. R. (1980) *Science* **210,** 545–547.
23. Selander, R. K., Musser, J. M., Caugant, D. A., Gilmour, M. N. & Whittam, T. S. (1987) *Microb. Pathog.* **3,** 1–7.
24. Sibley, L. D. & Boothroyd, J. C. (1992) *Nature (London)* **359,** 82–85.