# Clonal population structure and genetic diversity of *Candida albicans* in AIDS patients from Abidjan (Côte d'Ivoire)

François Nébavi*, Francisco J. Ayala†‡, François Renaud§, Sébastien Bertout*, Serge Eholié¶, Koné Moussa‖, Michèle Mallié*, and Thierry de Meeûs†‖**

*Laboratoire de Parasitologie et Mycologie Médicale, EA 2413, Faculté de Pharmacie, 34060 Montpellier Cedex 1, France; §Génétique et Evolution des Maladies Infectieuses, Equipe Evolution des Systèmes Symbiotiques, Unité Mixte de Recherche 2724, Institute de Recherche pour le Développement, Centre National de la Recherche Scientifique, BP 64501, 34394 Montpellier Cedex 5, France; ¶Service des Maladies Infectieuses et Tropicales, Centre Hospitalier Universitaire de Treichville, 01 BP V 3 Abidjan, Côte d'Ivoire; ‖Laboratoire de Parasitologie Mycologie, Institut Pasteur de Côte d'Ivoire, BP V 116 Abidjan, Côte d'Ivoire; and †Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697

Contributed by Francisco J. Ayala, December 30, 2005

**We have investigated the genotype at 14 enzyme-encoding loci in 275 isolates of the pathogenic yeast *Candida albicans* sampled from 42 HIV-positive patients (all but one with AIDS) from Abidjan (Côte d'Ivoire). We separately analyzed the following variables: patient, residence, age, gender, T cell count, hospitalization (yes or no), drug treatment, date of sampling, multilocus genotype, and serotype. The most important factors contributing to the genetic variability of *C. albicans* are individual patient and gender. Our data manifest that the population size of the parasite is relatively small within each patient, although larger in women than in men, and that, at least for the patients involved in the study, the transmission rate of *C. albicans* between human adults is very low. Most important is the inference that the prevailing mode of reproduction of *C. albicans* in natural populations is clonal, so that sexual reproduction is extremely rare, if it occurs at all.**

clonality | molecular epidemiology | population genetics | candidiasis | asexual reproduction

*Candida albicans* is a diploid opportunistic fungal pathogen present in the gastrointestinal and genitourinary flora of most healthy humans and other mammals (1, 2). In immunocompromised patients, *C. albicans* may invade host tissues, so that HIV patients frequently suffer from recurring oral candidiasis, which by itself is not life threatening; but, if the parasite gains access to the blood stream, it may cause severe damage in the kidneys, heart, or brain with fatal consequences (1–5). Epidemiological surveys manifest the dominance of this species in nosocomial infections caused by fungal organisms (6–8). Despite numerous studies and recent advances in the molecular genetics of this organism (for review, see refs. 2–4), its population biology remains largely unsettled with respect to such parameters as population size, transmission rate, and reproductive strategy (clonality versus sexuality), which are of considerable epidemiological and medical consequence (9–14).

Sampling design is decisive for ascertaining the population structure of natural populations of parasites, as well as other parameters of interest, especially so in clonal or partially clonal organisms (15–20). We present herein a population–genetics investigation of *C. albicans* in 42 candidiasis patients, all HIV-positive, which includes 5–10 isolates per patient; takes into account such significant variables as gender, age, locality, T cell count, drug treatment, relapse, and hospitalization; and seeks to determine the mode of reproduction of the parasite and the contribution of each variable to the epidemiology of the disease.

## Results

We have analyzed the diploid genotype at 14 enzyme-encoding loci in a total of 275 isolates of *C. albicans* from 42 patients, of which 13 were sampled twice, at day 0 [week (W) 1] and day 15 (W3); five *C. albicans* isolates were obtained from each patient at each time. The variables considered are as follows (Table 1): patient (1–42), residence (Abidjan or suburb), age, gender, T cell count (CD4), hospitalization (yes or no), drug treatment (amphotericin B, nystatin, or ketoconazole), date of sampling, multilocus genotype (G1–G37), and serotype (A or B). We found a total of 37 multilocus genotypes (Table 2). More than one multilocus genotype was found among the five isolates in eight cases (Tables 1 and 3): patients 3 (on W1 and W3), 7, 10, 27, 33, 36, and 38, which provided the data for estimating linkage disequilibrium among loci. Two serotypes, A and B, were found in patients 7 and 10; all other patients exhibit only one serotype, either A or B (Table 1). Twenty-nine patients were cured at the time of the second sampling, on W3; the multilocus genotypes and the serotypes of the other 13 patients on W3 are shown in Table 1. All patients were cured by the third sampling on W5.

The most significant factor contributing to the $F$ statistics is the individual patient. At W1, $F_{\text{Patient}} = 0.50$ ($P = 0.001$); for gender, $F_{\text{Sex}} = 0.03$ ($P = 0.031$). For relapsing patients, there is considerable differentiation between W1 andW3 of the same patient: $F_{\text{W1–W3}} = 0.20$ ($P = 0.001$). Separate analyses of multilocus genotype and serotype confirm these results. For genotype, $F_{\text{Patient}} = 0.92$ ($P = 0.001$), $F_{\text{Sex}} = 0.02$ ($P = 0.075$), and $F_{\text{W1–W3}} = 0.82$ ($P = 0.001$). For serotype, $F_{\text{Patient}} = 0.94$ ($P = 0.001$), $F_{\text{Sex}} = 0.08$ ($P = 0.078$), and $F_{\text{W1–W3}} = 1$ ($P = 0.017$). $F$ statistics corresponding to residence and date of sampling are never significant and are, therefore, excluded from subsequent analyses.

There is considerable heterozygote excess in the whole sample [$F_{\text{is}} = -0.85$ (where $F_{\text{is}}$ measures deviations from panmixia within each subpopulation), $P = 0.001$; $H_{\text{OBS}} = 0.259$, $H_{\text{EXP}} = 0.140$ (where $H_{\text{EXP}}$ is the average expected heterozygosity within each set of samples)]. Male samples display greater heterozygote excess ($F_{\text{is}} = -0.97$; $H_{\text{OBS}} = 0.295$, $H_{\text{EXP}} = 0.150$) than female samples ($F_{\text{is}} = -0.65$; $H_{\text{OBS}} = 0.208$, $H_{\text{EXP}} = 0.126$). The difference is highly significant ($P = 0.002$). There is heterozygote excess at each of the 10 loci that are polymorphic in males, but only at 8 of the 14 loci that are polymorphic in females. Thus, the SE of $F_{\text{is}}$ across loci is higher in females than in males (SE = 0.10 versus 0.01; $P = 0.046$). Null alleles might account for this difference. If all loci homozygous in all sets of female or male samples are removed, as well as those that are heterozygous in only one set [*Mdh2*, *G6pd*, *Hk1*, *Hk2*, *Fk*, *Gpi*, and *Mpi* (for explanations and accession numbers of the loci, see *Materials and Methods*)], $F_{\text{is}} = -0.75$, SE = 0.048 for females and $F_{\text{is}} = -0.96$, SE = 0.01 for males, differences that are statistically

Conflict of interest statement: No conflicts declared.

Abbreviation: W, week.

‡To whom correspondence may be addressed at: Department of Ecology and Evolutionary Biology, University of California, 321 Steinhaus Hall, Irvine, CA 92697-2525. E-mail: fjayala@uci.edu.

**To whom correspondence may be addressed. E-mail: demeeus@mpl.ird.fr.

**Table 1. Characteristics of the 42 patients sampled for *C. albicans***

| Patient* | Residence† | Age | Gender | CD4‡ | Hospitalization | Drug§ | Date,¶ day/month | Genotype (no., serotype)‖ |
|---|---|---|---|---|---|---|---|---|
| 1 | Abidjan | 32 | M | 27 | No | NYS | 07/05 | *G1* (5, A) |
| 2 | Abidjan | 28 | M | 102 | No | AMB | 12/05 | *G2* (5, A) |
| **3** | Abidjan | 51 | M | 87 | No | AMB | 26/05 | G3 (4, A), *G4* (1, A) |
| | | | | | | | 09/06 | G3 (3, A), *G4* (2, A) |
| **4** | Abidjan | 33 | F | 8 | Yes | NYS | 27/05 | *G5* (5, A) |
| | | | | | | | 10/06 | *G5* (5, A) |
| **5** | Abidjan | 18 | M | 15 | No | KTZ | 03/06 | G3 (5, A) |
| | | | | | | | 17/06 | G3 (5, A) |
| **6** | Suburb | 21 | M | 11 | Yes | AMB | 03/06 | G3 (5, A) |
| | | | | | | | 17/06 | G3 (5, A) |
| 7 | Abidjan | 30 | F | 38 | Yes | AMB | 08/06 | *G6* (4, A), G7 (1, B) |
| 8 | Abidjan | 20 | M | 22 | Yes | AMB | 21/06 | G3 (5, A) |
| 9 | Abidjan | 35 | M | 32 | Yes | KTZ | 29/06 | G8 (5, A) |
| 10 | Abidjan | 42 | F | 44 | No | KTZ | 29/06 | *G9* (1, A), *G10* (1, A), *G11* (2, B), *G12* (1, A) |
| **11** | Suburb | 19 | M | 221 | No | KTZ | 05/07 | *G13* (5, A) |
| | | | | | | | 19/07 | *G14* (5, A) |
| 12 | Abidjan | 19 | M | 25 | Yes | NYS | 08/07 | G15 (5, B) |
| **13** | Abidjan | 27 | M | 101 | Yes | KTZ | 13/07 | *G16* (5, A) |
| | | | | | | | 27/07 | *G16* (5, A) |
| **14** | Abidjan | 34 | M | 22 | No | NYS | 14/07 | G3 (5, A) |
| | | | | | | | 28/07 | *G17* (5, A) |
| 15 | Abidjan | 31 | F | 16 | Yes | AMB | 16/07 | G7 (5, B) |
| 16 | Suburb | 39 | M | 27 | No | KTZ | 16/07 | G8 (5, A) |
| 17 | Abidjan | 25 | F | 47 | No | NYS | 16/07 | G18 (5, A) |
| 18 | Abidjan | 26 | F | 17 | No | AMB | 19/07 | G19 (5, A) |
| **19** | Abidjan | 30 | M | 8 | Yes | AMB | 22/07 | G7 (5, B) |
| | | | | | | | 05/08 | G7 (5, B) |
| 20 | Suburb | 27 | M | 8 | Yes | AMB | 24/07 | *G20* (5, A) |
| 21 | Abidjan | 37 | F | 31 | No | KTZ | 26/07 | *G21* (5, A) |
| 22 | Suburb | 41 | F | 71 | No | KTZ | 30/07 | *G22* (5, A) |
| **23** | Suburb | 52 | M | 44 | Yes | AMB | 31/07 | G23 (5, A) |
| | | | | | | | 14/08 | G23 (5, A) |
| **24** | Abidjan | 26 | M | 22 | Yes | AMB | 04/08 | G23 (5, A) |
| | | | | | | | 18/08 | *G24* (5, A) |
| 25 | Abidjan | 33 | M | 9 | No | NYS | 10/08 | *G25* (5, A) |
| 26 | Abidjan | 25 | F | 12 | No | KTZ | 12/08 | G15 (5, B) |
| **27** | Abidjan | 19 | F | 44 | No | NYS | 17/08 | *G26* (5, A) |
| | | | | | | | 31/08 | G18 (2, B), G19 (3, B) |
| **28** | Abidjan | 21 | F | 11 | Yes | NYS | 20/08 | G15 (5, A) |
| | | | | | | | 03/09 | G15 (5, A) |
| 29 | Abidjan | 21 | M | 18 | Yes | KTZ | 22/08 | G15 (5, A) |
| 30 | Suburb | 37 | M | 44 | Yes | KTZ | 08/06 | G27 (5, A) |
| 31 | Abidjan | 40 | M | 27 | No | KTZ | 10/06 | *G28* (5, A) |
| **32** | Abidjan | 25 | F | 38 | Yes | KTZ | 19/06 | *G29* (5, B) |
| | | | | | | | 03/07 | *G29* (5, B) |
| 33 | Abidjan | 28 | F | 51 | No | KTZ | 03/07 | G15 (4, B), G27 (1, B) |
| 34 | Abidjan | 31 | F | 33 | No | KTZ | 11/07 | G30 (5, B) |
| 35 | Abidjan | 18 | M | 22 | No | KTZ | 11/07 | *G31* (5, A) |
| 36 | Suburb | 16 | F | 22 | Yes | KTZ | 14/07 | *G32* (3, A), *G33* (2, A) |
| 37 | Abidjan | 21 | F | 11 | Yes | KTZ | 06/08 | G34 (5, A) |
| **38** | Abidjan | 33 | M | 9 | Yes | KTZ | 12/08 | G34 (1, B), *G35* (3, B), *G36* (1, B) |
| 39 | Abidjan | 19 | M | 44 | No | KTZ | 20/08 | G34 (5, A) |
| 40 | Suburb | 22 | F | 285 | No | KTZ | 20/08 | G34 (5, A) |
| 41 | Abidjan | 32 | F | 46 | Yes | KTZ | 24/08 | G30 (5, A) |
| 42 | Abidjan | 43 | F | 8 | No | KTZ | 27/08 | *G37* (5, B) |

*Relapsing patients are in boldface type.
†Residence: City (Abidjan) or periphery (Suburb).
‡CD4 is T cell count per $mm^3$ of blood.
§AMB, amphotericin B; NYS, nystatin; KTZ, ketoconazole.
¶Date: 1997 for patients 1–29; 1998 for patients 30–42.
‖The 37 multilocus genotypes are given in Table 2; data in parentheses indicate the number of samples with a given genotype and serotype A or B. Multilocus genotypes in italics were found in only one patient.

**Table 2. Multilocus genotypes (G1–G37) of *C. albicans* at 14 enzyme-coding loci**

| Genotype | *Aat* | *Fk* | *Gpi* | *G6pd* | *Hk1* | *Hk2* | *Mdh1* | *Mdh2* | *Mpi* | *Np* | *Pep1* | *Pep2* | *Pep3* | *6Pgd* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G1 | 11 | 44 | 12 | 12 | 12 | 33 | 11 | 22 | 13 | 11 | 11 | 11 | 22 | 22 |
| G2 | 11 | 44 | 11 | 11 | 12 | 33 | 22 | 12 | 33 | 11 | 11 | 22 | 12 | 22 |
| G3 | 11 | 13 | 12 | 12 | 12 | 13 | 11 | 22 | 13 | 11 | 11 | 11 | 22 | 22 |
| G4 | 11 | 11 | 12 | 12 | 22 | 11 | 11 | 22 | 13 | 11 | 11 | 11 | 22 | 22 |
| G5 | 11 | 13 | 11 | 11 | 23 | 13 | 11 | 22 | 44 | 11 | 11 | 22 | 11 | 22 |
| G6 | 11 | 11 | 12 | 12 | 12 | 11 | 11 | 22 | 13 | 11 | 11 | 11 | 22 | 22 |
| G7 | 11 | 23 | 11 | 11 | 11 | 23 | 11 | 22 | 34 | 11 | 11 | 11 | 22 | 22 |
| G8 | 11 | 44 | 12 | 12 | 12 | 33 | 11 | 22 | 13 | 11 | 11 | 11 | 33 | 22 |
| G9 | 12 | 23 | 11 | 11 | 11 | 23 | 22 | 22 | 34 | 12 | 12 | 12 | 22 | 22 |
| G10 | 11 | 44 | 12 | 12 | 22 | 33 | 11 | 22 | 33 | 11 | 11 | 11 | 11 | 22 |
| G11 | 11 | 23 | 11 | 11 | 11 | 13 | 11 | 22 | 34 | 11 | 11 | 11 | 22 | 22 |
| G12 | 11 | 33 | 12 | 12 | 22 | 11 | 11 | 22 | 33 | 11 | 11 | 11 | 11 | 22 |
| G13 | 11 | 13 | 22 | 22 | 12 | 13 | 11 | 22 | 13 | 11 | 11 | 11 | 22 | 22 |
| G14 | 11 | 13 | 11 | 22 | 12 | 13 | 11 | 22 | 23 | 11 | 11 | 11 | 22 | 22 |
| G15 | 11 | 23 | 11 | 11 | 11 | 23 | 11 | 22 | 44 | 11 | 11 | 11 | 22 | 22 |
| G16 | 11 | 13 | 11 | 11 | 12 | 13 | 11 | 22 | 13 | 11 | 11 | 11 | 22 | 22 |
| G17 | 11 | 13 | 11 | 12 | 12 | 13 | 11 | 22 | 23 | 11 | 11 | 11 | 22 | 22 |
| G18 | 11 | 44 | 11 | 11 | 11 | 33 | 11 | 22 | 34 | 11 | 11 | 11 | 22 | 22 |
| G19 | 11 | 44 | 11 | 11 | 12 | 33 | 11 | 22 | 44 | 11 | 11 | 11 | 11 | 22 |
| G20 | 11 | 44 | 11 | 11 | 12 | 33 | 11 | 22 | 33 | 11 | 11 | 11 | 11 | 22 |
| G21 | 11 | 13 | 12 | 12 | 11 | 13 | 11 | 22 | 13 | 11 | 11 | 11 | 22 | 22 |
| G22 | 11 | 44 | 12 | 12 | 11 | 33 | 11 | 22 | 13 | 11 | 11 | 11 | 22 | 22 |
| G23 | 11 | 44 | 11 | 12 | 11 | 33 | 11 | 22 | 23 | 11 | 11 | 12 | 22 | 22 |
| G24 | 11 | 13 | 11 | 12 | 12 | 13 | 11 | 22 | 13 | 11 | 11 | 11 | 33 | 22 |
| G25 | 11 | 44 | 11 | 11 | 12 | 33 | 12 | 12 | 33 | 11 | 11 | 12 | 11 | 22 |
| G26 | 11 | 13 | 11 | 12 | 12 | 13 | 11 | 22 | 13 | 11 | 11 | 11 | 22 | 22 |
| G27 | 11 | 13 | 11 | 12 | 12 | 13 | 12 | 22 | 12 | 11 | 11 | 11 | 33 | 22 |
| G28 | 11 | 13 | 11 | 11 | 12 | 13 | 12 | 22 | 34 | 11 | 11 | 11 | 33 | 22 |
| G29 | 11 | 23 | 11 | 11 | 11 | 23 | 11 | 22 | 22 | 11 | 11 | 11 | 22 | 22 |
| G30 | 11 | 23 | 11 | 11 | 11 | 23 | 11 | 22 | 24 | 11 | 11 | 11 | 22 | 22 |
| G31 | 11 | 44 | 11 | 12 | 11 | 33 | 11 | 22 | 44 | 11 | 11 | 11 | 22 | 22 |
| G32 | 11 | 44 | 11 | 11 | 12 | 33 | 22 | 12 | 22 | 11 | 11 | 11 | 11 | 22 |
| G33 | 11 | 33 | 11 | 11 | 12 | 23 | 22 | 12 | 22 | 11 | 11 | 11 | 11 | 22 |
| G34 | 11 | 13 | 11 | 12 | 12 | 13 | 11 | 22 | 12 | 11 | 11 | 11 | 22 | 22 |
| G35 | 11 | 13 | 11 | 11 | 12 | 23 | 11 | 22 | 24 | 11 | 11 | 11 | 22 | 22 |
| G36 | 11 | 13 | 11 | 11 | 11 | 23 | 11 | 22 | 24 | 11 | 11 | 11 | 22 | 22 |
| G37 | 11 | 23 | 11 | 11 | 11 | 23 | 11 | 22 | 24 | 11 | 11 | 11 | 22 | 12 |

The numbers 1, 2, 3, and 4 represent different alleles at a given locus. The mean number of alleles per locus is 2.5.

significant ($P = 0.006$ and 0.008 for $F_{is}$ and SE, respectively). No $F_{is}$ difference exists between nonrelapsing and relapsing patients ($P = 0.396$ and 0.946 for females and males, respectively).

Linkage disequilibrium was analyzed in the eight sets of samples with more than one multilocus genotype (Table 3). There are 21 pairs of loci among the seven polymorphic loci (those with the frequency of the most common allele below 0.90), among which the overall samples $G$ statistic is significant between six pairs of loci. All loci except *Gpi* are involved in at least one significantly linked pair of loci. Linkage disequilibrium values, as measured by $R^2$ (21, 22), averaged over all loci pairs, are given in Table 3. For the multilocus statistics, six tests (of the eight possible ones) are significant (Table 3). The values of $\bar{r}_D$ measuring multilocus linkage are very high and significant in six of the eight sets of samples (Table 3). There is strong linkage disequilibrium between loci, as would be expected if the population structure of *C. albicans* is predominantly, or even exclusively, clonal.

**Table 3. Linkage disequilibrium in *C. albicans***

| Patient | Week | Gender | $\bar{R}^2$ | $G$ | $\bar{r}_D$ |
|---|---|---|---|---|---|
| 3 | 1 | M | 1 | 0.4* | 1* |
| 3 | 3 | M | 1 | 0.6** | 1** |
| 7 | 1 | F | 0.833 | 0.4*** | 1*** |
| 10 | 1 | F | 0.426 | 0.9*** | 0.291*** |
| 27 | 3 | F | 1 | 0.6** | 1** |
| 33 | 1 | F | 1 | 0.4*** | 1*** |
| 36 | 1 | F | 1 | 0.6 | 1 |
| 38 | 1 | M | 0.229 | 0.7 | 0.375 |

$R^2$ (21) is averaged over all loci. $G$ measures multilocus genotypic diversity, and $\bar{r}_D$ estimates multilocus linkage disequilibrium. *, $P = 0.05$; **, $P = 0.01$; ***, $P = 0.001$. $G$ and $\bar{r}_D$ for patients 36 and 38 are not significant.

No difference in linkage disequilibrium measures are statistically significant between isolates sampled in females and in males.

We have assessed whether *C. albicans* is strictly clonal according to the criterion proposed by de Meeûs and Balloux (19). In a strictly clonal population, strongly subdivided into numerous demes, $F_{st}$ (a measure of differentiation between subpopulations in the total population) should not be significantly different from $F_{st}' = -F_{is}/(1 - F_{is})$ and $F_{it} = 0$ (where $F_{it}$ measures inbreeding resulting from both nonrandom union of gametes within subpopulations and from population structure). This is the case, as shown in Fig. 1, under the reasonable assumption that numerous individuals carry *C. albicans* in Abidjan. Under this assumption, the number of migrants exchanged between populations (individual patients) per yeast generation is the product $Nm$ (where $N$ is the *C. albicans* population size in an individual, and $m$ is the per capita transmission

EVOLUTION

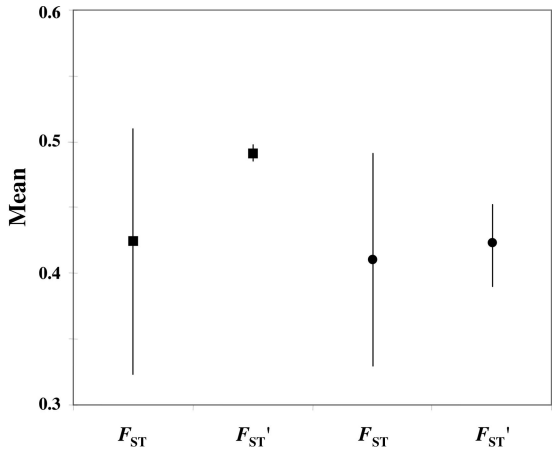


**Fig. 1.** Comparison between $F_{st}$ and $F_{st}' = -F_{is}/(1 - F_{is})$ in males (■) and females (●). The bars mark 95% confidence intervals obtained by bootstrapping over loci. $F_{it}$ values (data not shown) are not significantly above 0 ($P = 0.767$ and $P = 1$ for females and males, respectively) on the basis of 10,000 random permutations. Values and tests were performed on W1 samples (before treatment) with the seven most polymorphic loci defined in *Results*.

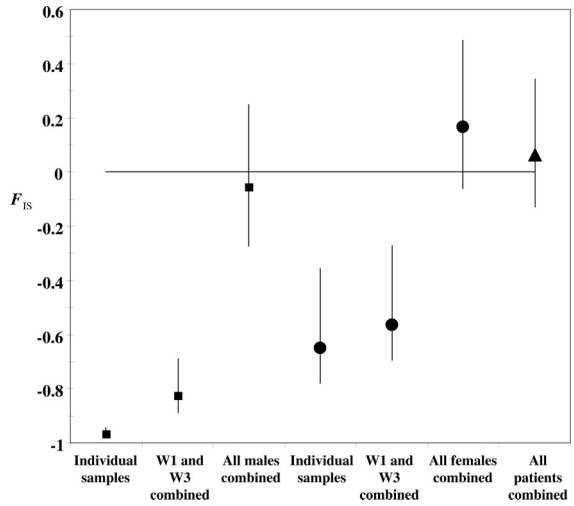


**Fig. 2.** Population structure of *C. albicans* in AIDS patients from Abidjan (pooling method in ref. 23). Squares, male samples; circles, female samples; triangles, males and females combined. If the samples pooled belong to the same reproductive unit (deme), no change in the $F_{is}$ estimate is expected. $F_{is}$ increases when populations with different genetic composition are combined; this is the case when samples from W1 and W3 are combined for males and when all males or all females are combined. Bars are 95% confidence intervals obtained by bootstrapping over loci (24).

rate between individuals), which can be estimated by $Nm = -(1 + F_{is})/4F_{is}$ (ref. 19). For our data, $Nm = 0.09$ for isolates from female patients, and $Nm = 0.01$ for isolates from male patients. This 9-fold difference could be due to a larger *C. albicans* population in women, to a greater rate of transmission between women than between men, or to a combination of both.

## Discussion

Our investigation yields five significant results: (*i*) the individual patient is the factor that contributes most to genetic diversity in *C. albicans* populations; (*ii*) for the patients involved in our study, there is very low transmission rate of *C. albicans* between adult individuals; (*iii*) the population size of the parasite is relatively small within each patient; (*iv*) the population characteristics of *C. albicans* differ between male and female patients; and (*v*) in the population investigated, the prevailing mode of reproduction of *C. albicans* is clonal, so that sexual reproduction occurs, if at all, with a frequency smaller than $10^{-5}$. We will discuss these results in turn.

The *C. albicans* strains found among the 42 patients vary from individual to individual, so that the genetic makeup of *C. albicans* is rarely identical in two or more patients. Of the 37 multilocus genotypes, 27 (72%) are found only once among the 42 patients (Table 1).

The extremely low values of *Nm*, 0.01 in males and 0.09 in females, indicate that both the clonal effective population size (*N*) of *C. albicans* in each patient and the rate of transmission (*m*) between patients are low. The difference in the *Nm* value between males and females indicates that either *N* or *m* or a combination of the two is ≈10 times greater in females than males. Indeed, a likely possibility is that *C. albicans* population size is generally larger in women than in men. If we assume that *m* is approximately the same for the two genders, *N* would be ≈10 times larger in women than in men.

A small *C. albicans* population size is consistent with the genetic differentiation observed between W1 and W3 isolates of the 13 patients (four females and nine males) that were sampled twice. These differences may have come about by genetic drift, although the possibility of natural selection in response to the drug treatments cannot be excluded. In any case, drug treatment contributes to population size reduction of *C. albicans* in individual patients; indeed, candidiasis is absent by W3 in many individuals, and, by W5, no *C. albicans* are found in any patient. The $F_{is}$ values indicate that the genotype differences between W1 and W3 are greater for males than for females ($F_{is}$ is significantly greater for the W1 and W3 samples combined than separately in the case of males, but not in the case of females) (Fig. 2). This gender difference may also be due to a larger *C. albicans* population in females.

The low value of *Nm* favors overall genetic diversity of *C. albicans* in the human population, as well as genetic differentiation between patients. Such a population structure calls for appropriate sampling when seeking to determine the population structure of *C. albicans* or the suitability of one or another medical treatment. If we had obtained only one isolate per patient, as it is often done (25–33), we would have failed to discover the presence of more than one strain in each of the eight samples in which we found two or more. When strain variation among patients is ignored, the observations may be uninterpretable (34–35) or unreliable, as in the apparent differences between invasive and noninvasive strains of *C. albicans* (36) or between the strains found in patients with different religious beliefs (37).

Sampling repeatedly from the same individual as in some other studies (38, 39) is necessary but not sufficient. As can be seen here, measuring between patients' genetic differentiation is also important. Combining the samples from different individuals, even within the same gender, may yield spurious results, as is apparent in Fig. 2, in which the deficiency of heterozygotes disappears when the male samples or the female samples, or all are combined; that is, mean $F_{is}$ approaches 0, the value expected under panmixia.

The degree of *C. albicans* genetic differentiation between men and women is relatively small ($F_{Sex} = 0.03$) compared with the differentiation among human individuals of either gender (Fig. 1). There is, however, a gender difference in overall heterozygosity, which is greater in males (significantly so when individual samples are compared between males and females) (Fig. 2). Differences in immune competence between males and females are common in vertebrates (40), but we have not observed significant gender differences with respect to CD4 count or probability of relapsing (generalized linear model; $P > 0.20$). The possibility remains, however, that some of the differences observed in levels of heterozygosity, and perhaps others, may result from immunological and/or behavioral differences.

The nearly complete and highly significant linkage disequilibrium in our dataset suggests that the mode of reproduction of *C. albicans* is completely, or nearly so, clonal (Table 3). This is confirmed by the lack of significant difference between $F_{st}$ and $F_{st}'$ (Fig. 1).

The extent of clonal rather than sexual reproduction in *C. albicans* has been a subject of controversy (e.g., refs. 41 and 42), with limited evidence that it might predominantly reproduce clonally rather than sexually (14, 25, 26). There can be little doubt that *C. albicans* has the capacity to reproduce sexually (see below) and that it reproduces sexually in the laboratory (1, 43–45). However, the issue is whether it reproduces sexually in natural populations and to what extent. Some authors have claimed that genetic recombination readily occurs in natural populations of *C. albicans* (27, 32, 46, 47), but their evidence is far from conclusive because it relies mostly on the lack of statistical significance of tests for segregation or recombination or in the observation of all possible allele combinations between pairs of loci. Yet, absence of statistical significance is not proof of the null hypothesis, particularly when tests have low statistical power or are performed on unsuitable samples, as is the case when data for different patients, different dates, or different locations are combined. Indeed, for the eight subsamples with multiple genotypes, the average $R^2 = 0.51$ (Table 3), which indicates strong linkage disequilibrium, but if all 55 subsamples (42 patients, 13 of them sampled twice) are combined into one, $R^2 = 0.09$. The misleading consequence of combining separate samples is also apparent in Fig. 2. $F_{is}$ approaches $-1$ in males, but when all male or all female samples are combined, or both of them, $F_{is}$ confidence intervals include 0. There is not, at present, definitive evidence of sexual reproduction of *C. albicans* in nature. Nevertheless, the higher mean and variance $F_{is}$ in females than in males observed in our data (Fig. 2) may be due to the occasional occurrence of sexual reproduction (15–18). Some of the genotypes observed in female patients 7, 10, 27, 33, and 36 may have arisen by genetic recombination in the patients themselves or in ancestral strains.

*C. albicans* possesses a single mating-type locus (*MTL*) on chromosome 5, which is heterozygous ($a/\alpha$) in 97% of clinical isolates (2, 48, 49). Mating can only occur between strains that are homozygous for one or the other allele ($a/a$ or $\alpha/\alpha$). Meiosis is extremely rare or totally absent, because it has never been observed in *C. albicans* (44). Nevertheless, homozygosis occasionally occurs as a consequence of chromosome loss or, more rarely, mitotic recombination (48). Genome rearrangements and aneuploidy are observed in clinical isolates of *C. albicans* (3, 50). In particular, fluconazol-resistant strains frequently exhibit chromosomal rearrangements (2). Moreover, stress conditions may induce nondisjunction, leading to homozygosity at the *MTL* and thus to the opportunity for sexual reproduction (43). Clinical isolates, including fluconazole-resistant strains, are occasionally homozygous at the *MTL* and thus mating competent (2). According to the theory of de Meeûs and Balloux (19), the seven most polymorphic loci of our data yield an estimate of $10^{-5}$ or, very probably, fewer *C. albicans* cells arisen by sexual reproduction. Given this rarity, it seems likely that the genetic signature of sexual reproduction, if it occurs, would be quickly lost, particularly in smaller populations, such as in male patients.

In conclusion, our study of HIV-positive candidiasis patients confirms that, in the natural populations that we have investigated, *C. albicans* reproduces clonally, with extremely rare exceptions, if any. Additional studies will be needed to confirm or not the generality of this conclusion. We have also shown that, to ascertain the population structure of this common human pathogen, human patients should each be sampled repeatedly, that the analysis of the genetic data should consider each patient separately, and that male and female patients should also be separately analyzed. One significant observation is that the rate of transmission of *C. albicans* between the adults involved in this study is extremely low and excludes nosocomial infection as a common event among the patients.

## Materials and Methods

**Samples.** *C. albicans* isolates were obtained from 42 HIV-positive patients with oral candidiasis from Abidjan and suburbs (Table 1), all with AIDS, except patient 40. Each sample consisted of five *C. albicans* isolates, obtained by buccal swabbing, followed by suspension in 1 ml of distilled water and growth on Sabouraud chloramphenicol agar medium (BioMerieux, Marcy l'Etoile, France) at 37°C. Patients were randomly chosen with respect to antifungal treatment: 24 patients were treated with ketoconazole (200 mg twice daily), 10 patients with amphotericin B (500 mg in drinkable suspension four times daily), and 8 patients with nystatin ($10^5$ units four times daily, gynecological tablets). Three samplings were undertaken: all 42 patients before treatment [day 0 (W1)], 15 days after treatment (W3), and 30 days after treatment (W5). At W3, 29 patients were cured; the remaining 13 provided a second set of isolates (Table 1). At W5, all patients were cured.

**Enzyme Assays.** Samples were prepared as described previously (25). Starch-gel electrophoresis and enzyme assays were performed according to refs. 51 and 52 for 14 putative gene loci: aspartate-aminotransferase (EC 2.6.1.1; *Aat*), fructo-kinase (EC 2.7.1.4; *Fk*), glucose-phosphate-isomerase (EC 5.3.1.9; *Gpi*), glucose-6-phosphate-dehydrogenase (EC 1.1.1.49; *G6pdh*), hexokinase (EC 2.7.1.1; *Hk9* and *Hk2*), malate-dehydrogenase (EC 1.1.1.37; *Mdh1* and *Mdh2*), mannose-6-phosphate-isomerase (EC 5.3.1.8; *Mpi*), purine-nucleoside-phosphorylase (EC 2.4.2.1; *Np*), peptidase A (EC 3.4.13; substrate Val-Leu, *Pep1*), peptidase B (EC 3.4.13; substrate Leu-Gly-Gly, *Pep2*), peptidase D (EC 3.4.13; substrate Phe-Pro, *Pep3*), and 6-phosphogluconate-dehydrogenase (EC 1.1.1.43; *6Pgd*).

A total of 37 different multilocus genotypes (Table 2, G1–G37) were characterized among the 55 sets of samples, each consisting of five isolates.

**Data Analysis.** We performed two principal component analyses, for W1 and W3 samples, with the program PCA-GEN version 1.2 (J. Goudet, University of Lausanne, Lausanne, Switzerland), following the broken-stick model (53). The coordinates of the samples were retrieved and submitted to a generalized linear model analysis by using S-PLUS 2000 professional release 2 (Mathsoft, Cambridge, MA). The variables considered were residence, CD4 count, age, sex, hospitalization, relapse status (yes or no), sample date, and drug treatment (for W3 samples). Samples were considered as contemporaneous if separated ≤6 days. We dropped less significant terms in turn, by using the Akaike Information Criterion (S-PLUS 2000 Guide to Statistics). The significance of each factor was tested by an *F* test. Significant factors were used for a HIERFSTAT analysis (see below), with two additional factors: individual patient and W1 versus W3 for relapsing patients.

The generalized linear model analyses undertaken on the significant axes of principal component analysis from W1 samples (five axes) and W3 samples (three axes) lead to identify four significant factors: CD4 count (for axis 2 of W1), which is significant exclusively because of two outliers (patients 11 and 40 with CD4 counts of 221 and 285, respectively) and was disregarded, residence (for axis 5 of W1 and axis 3 of W3), gender (for axis 2 of W1 and axis 1 of W3), and date of sampling (for axis 1 of W1 and axis 2 of W3). Thus, only residence, gender, and date of sampling were kept and added to individual patient and week of sampling (W1 or W3) for the hierarchical *F* statistics analysis.

*F* statistics (54) measure departures from panmixia, so that $F_{is}$ measures deviations from panmixia within each subpopulation, $F_{st}$ measures differentiation between subpopulations in the total population, and $F_{it}$ measures inbreeding resulting from both nonrandom union of gametes within subpopulations and from population

structure. More than two levels may exist. HIERFSTAT version 0.03-2 (20) is a package for the statistical software R (55), which computes hierarchical $F$ statistics from any number of hierarchical levels (20). The significance of the hierarchical $F$ statistics was tested by 1,000 randomizations of each relevant factor among the entities belonging to the next level (e.g., isolates among patients), keeping them separate for other factors (e.g., gender). To implement such analyses with crossed factors such as gender and date of sampling, we estimated and tested each level in different files to control for the other effects. $F_{is}$ was tested by randomizing alleles among individuals within each of the 55 subsamples by FSTAT 2.9.4 (24).

Each multilocus genotype (Table 2, G1–G37) was treated as a different allelic state at a single locus (as in ref. 56). Serotypes A and B were also treated as two alleles, 1 and 2, at a single locus.

For comparisons of $F_{is}$ between different groups of samples (males versus females, relapsing versus nonrelapsing patients), 10,000 randomizations were made of the subsamples between the groups. The SE of $F_{is}$ was estimated by jackknifing over loci and, for each locus, by jackknifing over populations (24); the comparison between males and females was performed by a Wilcoxon's test, with locus as the pairing unit.

Linkage disequilibrium between pairs of loci was measured by $R^2$ (21, 22) and generalized for the multiallele case as in FSTAT 2.9.4. Significance of association between each pair of loci over all samples was determined by randomization by using the log-likelihood ratio $G$ statistic (57), with Bonferroni's correction. Only "polymorphic" loci (no allele with a frequency $\geq 0.90$) were considered. Multilocus linkage disequilibrium was implemented by MULTILOCUS 1.3b (58). As recommended previously (16, 18), only the multilocus genotypic diversity $G$ and multilocus standardized linkage disequilibrium $\bar{r}_D$ (58) were used for this analysis, with significance obtained by randomization (59). This test was made within each subsample.

We thank D. Castel for technical assistance, J. Cilag for providing ketoconazole, and Jerôme Goudet for his help. T.d.M. and F.R. are supported by the Centre National de la Recherche Scientific and the Institut de Recherche pour le Développement.

1. Hull, C. M., Raisner, R. M. & Johnson, A. D. (2000) *Science* **289,** 307–310.
2. Berman, J. & Sudbery, P. E. (2002) *Nat. Rev. Genet.* **3,** 918–928.
3. De Backer, M. D., Magee, P. T. & Pla, J. (2000) *Annu. Rev. Microbiol.* **54,** 463–498.
4. Bougnoux, M. E., Aanensen, D. M., Morand, S., Théraud, M., Spratt, B. G. & d'Enfert, C. (2004) *Infect. Genet. Evol.* **4**, 243–252.
5. Beck-Sague, C. & Jarvis W. R. (1993) *J. Infect. Dis.* **167,** 1247–1251.
6. Verduyn Lunel, F., Meis, J. F. G. M. & Voss, A. (1999) *Diagn. Microbiol. Infect. Dis.* **34,** 213–220.
7. Gupta, N., Haque, A., Lattif, A. A., Narayan, R. P., Mukhopadhyay, G. & Prasad, R. (2004) *Mycopathologia* **158,** 397–405.
8. Correira, A., Sampaio, P., Almeida, J. & Pais, C. (2004) *J. Clin. Microbiol.* **42,** 5899–5903.
9. Milgroom, M. G. (1996) *Annu. Rev. Phytopathol.* **34,** 457–477.
10. Tibayrenc, M. (1998) *Int. J. Parasitol.* **28,** 85–104.
11. Tibayrenc, M. (1999) *Annu. Rev. Genet.* **33,** 449–477.
12. Taylor, J. W., Geiser, D. M., Burt, A. & Koufopanou, V. (1999) *Clin. Microbiol. Rev.* **12,** 126–146.
13. Tibayrenc, M., Kjellberg, J. & Ayala, F. J. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2414–2418.
14. Tibayrenc, M. & Ayala, F. J. (2002) *Trends Parasitol.* **18,** 405–410.
15. Halkett, F., Simon, J. F. & Balloux, F. (2005) *Trends Ecol. Evol.* **20,** 194–201.
16. de Meeûs, T., Lehmann, L. & Balloux, F. (2006) *Infect. Genet. Evol.*, in press.
17. Balloux, F., Lehmann, L. & de Meeûs, T. (2003) *Genetics* **164,** 1635–1644.
18. de Meeûs, T. & Balloux, F. (2004) *Infect. Genet. Evol.* **4,** 345–351.
19. de Meeûs, T. & Balloux, F. (2005) *Mol. Ecol.* **14,** 2695–2702.
20. Goudet, J. (2005) *Mol. Ecol. Notes* **5,** 184–186.
21. Weir, B. S. (1979) *Biometrics.* **35,** 235–254.
22. Weir, B. S. (1996) *Genetic Data Analysis II* (Sinauer, Sunderland, MA), 2nd Ed.
23. Goudet, J., de Meeûs, T., Day, A. J. & Gliddon, C. J. (1994) in *Genetics and Evolution of Aquatic Organisms*, ed. Beaumont, A. (Chapman and Hall, London), pp. 81–95.
24. Goudet, J. (1995) *J. Hered.* **86,** 485–486.
25. Pujol, C., Reynes, J., Renaud, F., Raymond, M., Tibayrenc, M., Ayala, F. J., Janbon, F., Mallié, M. & Bastide, J. M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 9456–9459.
26. Boerlin, P., Boerlin-Petzold, F., Goudet, J., Durussel, C., Pagani, J. L., Chave, J. P. & Bille, J. (1996) *J. Clin. Microbiol.* **34,** 1235–1248.
27. Gräser, Y., Volovsek, M., Arrington, J., Schonian, G., Presber, W., Mitchell, T. G. & Vilgalys, R. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 12473–12477.
28. Pfaller, M. A., Lockhart, S. R., Pujol, C., Swails-Wenger, J. A., Messer, S. A., Edmond, M. B., Jones, R. N., Wenzel, R. P. & Soll, D. R. (1998) *J. Clin. Microbiol.* **36,** 1518–1529.
29. Xu, J., Mitchell, T. G. & Vigalys, R. (1999) *Mol. Ecol.* **8,** 59–73.
30. Lott, T. J., Holloway, B. P., Logan, D. A., Fundyga, R. & Arnold, J. (1999) *Microbiology* **145,** 1137–1143.
31. Arnaviehle, S., de Meeûs, T., Blancart, A., Mallié, M., Renaud, F. & Bastide, J. M. (2000) *Mycoses* **43,** 109–117.
32. Fundyga, R. E., Lott, T. J. & Arnold, J. (2002) *Infect. Genet. Evol.* **2,** 57–68.
33. Bougnoux, M. E., Morand, S. & d'Enfert, C. (2002) *J. Clin. Microbiol.* **40**, 1290–1297.
34. Clemons, K. V., Feroze, F., Holmberg, K. & Stevens, D. A. (1997) *J. Clin. Microbiol.* **35,** 1332–1336.
35. Karahan, Z. C., Güriz, H., Ağirbaşli, H., Balaban, N., Gömen, J. S., Aysev, D. & Akar, N. (2004) *Mycoses* **47,** 465–469.
36. Pinto de Andrade, M., Schönian, G., Forche, A., Rosado, L., Costa, I., Müller, M., Presber, W., Mitchell, T. G. & Tietz, H. J. (2000) *Int. J. Med. Microbiol.* **290,** 97–104.
37. McCullough, M. J., Jorge, J. J., Lejbkowicz, F., Lefler, E., Nassar, F. Clemons, V. & Stevens, D. A. (2004) *Mycopathologia* **158,** 39–41.
38. Reynes, J., Pujol, C., Moreau, C., Mallié, M., Renaud, F., Janbon, F. & Bastide, J. M. (1996) *FEMS Microbiol. Lett.* **137,** 269–273.
39. Mata, A. L., Rosa, R. T., Rosa, E. A. R., Gonçalves, R. B. & Hölfing, J. F. (2000) *Oral Microbiol. Immunol.* **15,** 350–354.
40. Klein, S. L. (2000) *Behav. Processes* **51,** 149–166.
41. Tibayrenc, M. (1997) *Trends Microbiol.* **5,** 253–254.
42. Vilgalys, R., Gräser, Y. & Presber, W. (1997) *Trends Microbiol.* **5,** 254–256.
43. Magee, B. B. & Magee, P. T. (2000) *Science* **289,** 310–313.
44. Tzung, K. W., Williams, R. M., Scherer, S., Federspiel, N., Jones, T., Hansen, N., Bivolarevic, V., Huizar, L., Komp, C., Surzycki, R., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98,** 3249–3253.
45. Lockhart, S. R., Daniels, K. J., Zhao, R., Wessels, D. & Soll, D. R. (2003) *Eukaryotic Cell* **2,** 49–61.
46. Forche, A. Schönian, G., Gräser, Y., Vilgalys, R. & Mitchell, T. G. (1999) *Fungal Genet. Biol.* **28,** 107–125.
47. Tavanti, A., Gow, N. A. R., Maiden, M. C. J., Odds, F. C. & Duncan, J. S. (2004) *Fungal Genet. Biol.* **41,** 553–562.
48. Wu, W., Pujol, C., Lockhart, S. R. & Soll, D. R. (2005) *Genetics* **169,** 1311–1327.
49. Lockhart, S. R., Wu, W., Radke, J. B., Zhao, R. & Soll, D. R. (2005) *Genetics* **169,** 1883–1890.
50. Selmecki, A., Bergmann, S. & Berman, J. (2005) *Mol. Microbiol.* **55,** 1553–1565.
51. Pasteur, N., Pasteur, G., Bonhomme, F., Catalan, J. & Britton-Davidian, J. (1987) *Manuel Technique de Génétique par Electrophorèse des Protéïnes* (Lavoisier, Paris).
52. Richardson, B. J., Baverstock, P. R. & Adams, M. (1986) *Allozyme Electrophoresis: A Handbook for Animal Systematics and Population Studies* (Academic, Sydney).
53. Legendre, P. & Legendre, L. (1998) *Numerical Ecology* (Elsevier Science, Amsterdam), 2nd Ed.
54. Wright, S. (1965) *Evolution* **19,** 395–420.
55. R Development Core Team (2004) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna).
56. de Meeûs, T., Renaud, F., Mouveroux, E., Reynes, J., Galeazzi, G., Mallié, M. & Bastide, J. M. (2002) *J. Clin. Microbiol.* **40,** 2199–2206.
57. Sokal, R. R. & Rohlf, F. J. (1981) *Biometry* (Freeman, New York), 2nd Ed.
58. Agapow, P. M. & Burt, A. (2001) *Mol. Ecol. Notes* **1,** 101–102.
59. Burt, A., Carter, D. A., Koenig, G. L., White, T. J. & Taylor, J. W. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 770–773.