Molecular Epidemiology of the Global and Temporal Diversity of *Candida albicans*

Michael J. McCullough,1,2* Karl V. Clemons,1 and David A. Stevens1

The epidemiology of *Candida albicans* has changed with the rise in immunocompromised patients and the pressures of antifungal treatment and prophylaxis. We assessed the genotype distribution of recently obtained, globally diverse isolates in comparison with isolates recovered in the United States and United Kingdom before 1985, in order to determine temporal and geographic differences. We used *Eco*RI digestion of cellular DNA to generate restriction fragment length polymorphisms, dividing the isolates into 4 groups. From 15 diverse geographic areas, 439 isolates obtained over 20 years were divided into 121 genotypes within groups A (289 isolates), B (85), C (56), and D (9). Differences in genotype distribution existed among the localities (*P* < .0001) and between isolates obtained before 1990 versus those recovered since then (*P* = .009). Comparison of pre-1985 United States/United Kingdom isolates with post-1994 United States isolates revealed a trend toward a changing genotype distribution (*P* = .057). Global post-1985 isolates were different in genotype distribution from United States/United Kingdom isolates (*P* < .0001). The distribution of isolates from Israel was unique (*P* < .0001). These differences could be due in part to the increasing prevalence of group C strains worldwide.

The dimorphic yeast *Candida albicans* has been recognized as an increasingly important human pathogen, particularly in the host immunocompromised by advanced age, immunosuppressive therapy, or infection [1]. A variety of methods have been utilized to differentiate strains of *C. albicans* [2–4]. Recent advances in molecular biology have allowed restriction enzyme analysis of *C. albicans* genomic DNA, useful for epidemiological purposes [5–7]. One of the earliest reports of this methodology described the use of a DNA extraction protocol, followed by digestion with the restriction endonuclease *Eco*RI and electrophoresis of the resultant DNA fragments [8]. This method proved to be both simple and reproducible.

This DNA typing method was applied to assess the epidemiology of *C. albicans* isolated from diverse areas in the United States and the United Kingdom [9]. This previous study divided all strains studied into 2 broad genotype groups (A and B) on the basis of the position of a dimorphic band (3.7 or 4.2 kb, respectively) [9]: these bands are DNA-encoding ribosomal RNA (rDNA). The same methodology also has been utilized in several investigations that assessed strains of *C. albicans* isolated from heroin addicts in Glasgow [10] and Spain [11], as well as medical students in Singapore [12]. Reports on the latter 2 of these studies described 2 new genotype groups (C and D), established on the basis of the presence (genotype C) or the absence (genotype D) of both dimorphic bands [11, 12]. Further analyses [13] of the genotype group D isolates are presented in the Discussion.

Differences in geographic distribution of *C. albicans* genotypes were suggested in early studies [9] and have been corroborated [11, 12, 14, 15], although not always [16]. The aim of the current study was to assess the genotype of recently isolated, globally diverse strains of *C. albicans* for comparison with strains isolated from the United States and United Kingdom prior to 1985, in order to ascertain whether reported differences in genotypes among localities and patient populations were more widespread. We believe this is the largest molecular epidemiological survey of *Candida* undertaken.

Materials and Methods

Recent (1994 and later) isolates of *C. albicans* were obtained from France (courtesy of Dr. Francoise Dromer, Pasteur Institute,
Paris) \( (n = 21) \); Zimbabwe (courtesy of Drs. Val Robertson and David Katzenstein, University of Zimbabwe, Harare) \( (n = 19) \); Colombia (courtesy of Dr. Angela Restrepo, Corporacion para Investigaciones Biologicas, Medellin) \( (n = 20) \); Japan (courtesy of Dr. Katsuhiko Kamei, Research Center for Pathogenic Fungi and Microbial Toxicses, Chiba University, Chiba) \( (n = 52) \); Israel (courtesy of Dr. Eli LeRi, Western Galilee Hospital, Nahariya) \( (n = 30) \); California (California Institute for Medical Research, San Jose) \( (n = 17) \); Texas (courtesy of Dr. Michael Rinaldi, Department of Pathology, University of Texas, San Antonio) \( (n = 36) \); Iowa (courtesy of Dr. Michael Pfaller, Clinical Laboratory, University of Iowa, Iowa City) \( (n = 25) \); and Georgia (courtesy of Dr. Mike McNeil, Mycotic Disease Branch, Centers for Disease Control and Prevention, Atlanta) \( (n = 10) \).

Isolates made available from previously published studies, for comparative purposes, were initially isolated before 1982 in the United States and United Kingdom \( [9] (n = 91) \); in 1985 in Glasgow \( [10] (n = 13) \); from 1985 to 1990 in Spain \( [11] (n = 54) \); from 1990 to 1995 in Australia \( [17] (n = 23) \); and in 1995 in Singapore \( [12] (n = 23) \). Finally, 5 isolates initially isolated in the United Kingdom in 1986 were provided by Dr. Frank Odds of Janssen Pharmaceutica (Beerse, Belgium).

All isolates were tentatively identified by the API 20C system (bioMérieux, Marcy l’Etoile, France) as C. albicans. The isolates were sent to the California Institute for Medical Research (San Jose, CA) for further analysis. For short-term (several months) storage, the isolates were stored under water. Long-term storage of isolates was done by freezing at \(-80^\circ C\) in 40% glycerol.

Cellular DNA was isolated by means of methods described elsewhere \([8]\). Approximately 3 \( \mu g \) of the resultant DNA was digested with 40 U of the restriction endonuclease EcoRI overnight at \(37^\circ C\). The DNA fragments were separated through an agarose gel (0.8\% w/v; SeaKem Gold, FMC Bioproducts, Rockland, ME) in TAE buffer (40 mM of Tris-acetate and 0.2 mM of EDTA; pH 8.3) for 18 h at 2 V/cm and visualized by ultraviolet transillumination after ethidium-bromide staining. The reproducibility of this method, including insensitivity to serial passage of an isolate, identical analyses when done repeatedly with different blinded observers (a point confirmed in the present study), and reproducibility between laboratories, has been detailed in prior publications \([8â€“13, 18]\).

Digital images of the resultant DNA restriction fragment length polymorphisms (RFLPs) were captured with a charged-coupled device camera via the BioImage AQ gel documentation system (BioImage, Ann Arbor, MI). These images were analyzed by the BioImage AQ software, and the resultant band patterns were matched for percentage similarities by the Dice method with a 2% interband tolerance. Only those bands utilized for genotype subgrouping in the original description of this method \([8]\) were included in this analysis. Dendrograms were generated from this analysis by the unweighted pair-group method with arithmetic average.

As mentioned in the Introduction, sorting into broad genotype groups \( (A, B, C, \) and \( D) \) relied on the 3.7- and 4.2-kb bands \([8â€“13]\); sorting into subgroups \( (\text{types}) \) relied on the variable bright polymorphic bands in the 2-2.3-kb and 6-9-kb ranges \([8â€“12, 18]\).

From these analyses, isolates that had a percentage of similarity of >85% were considered to be possibly of the same genotypic subgroup \( (\text{type}) \), and their EcoRI-digested cellular DNA was electrophoresed side by side on the same agarose gel to confirm this subgrouping. The dendrogram analysis was then used only as an aid in the initial screening of isolates for further typing. Those isolates with <85% similarity by this methodology or whose banding patterns were shown to be different when run side by side were considered a unique subgroup \( (\text{type}) \).

Statistical analyses in this study included use of the \( \chi^2 \) test in contingency analysis, and differences among or between groups were assumed to be significant when the probability \( (P) \) was \( \leq .05 \).

Type strains of the original genotypic types described \([9, 10]\) have been deposited in the American Type Culture Collection on request, and full sets of them have been provided to investigators for epidemiological purposes. The newly delineated types are likewise available to investigators.

Results

Genotype groups. A total of 439 isolates of C. albicans obtained over a 20-year period were assessed from 15 geographically diverse areas (table 1). These isolates were divided, as has previously described \([9, 10, 12]\), into types. The isolates from each locality were initially compared with the “original” type strains—that is, those from the 2 earlier studies \([9, 10]\)—which were representative of each of the 33 genotypes observed in those 2 studies (genotype group \( A = 23 \) types and genotype group \( B = 10 \) types).

The new sites studied revealed 88 new genotypes: 43 in genotype group \( A, 25 \) in group \( B, 18 \) in group \( C, \) and \( 2 \) in group \( D \). Twenty-five of the 88 new types identified in this study were found in 2 or more of the new locations studied. We distinguished a total of 121 different types within genotype groups \( A \) (289 isolates; 66 types), \( B \) (85 isolates; 35 types), \( C \) (56 isolates; 18 types), and \( D \) (9 isolates; 2 types) (table 1).

The new group \( C \) types were no more focal than the group \( A \) or \( B \) types; several were found in different countries on different continents. The group \( B \) isolates in this study fell almost evenly between the original 10 group \( B \) types \([9, 10]\) and the 25 new group \( B \) types; >60% of the group \( A \) isolates studied fell into the original 23 group \( A \) types \([9, 10]\), and the remainder into the 43 new group \( A \) types. Appearance in multiple countries of the new isolates typed also tended to be more common for the 33 “original” types than for the 68 new group \( A \) and \( B \) types. These 2 findings suggest that the initial 33 types encountered, as a group, may be more globally distributed.

The isolates of genotype \( D \) \((n = 9)\) had only 2 EcoRI RFLP patterns (D1, 4 isolates; and D2, 5 isolates). These were similar to each other but very different from those of the remaining 430 isolates of C. albicans (figure 1). The 3.7- and 4.2-kb bands used to discriminate among groups \( A \) (3.7-kb band), \( B \) (4.2-kb band), and \( C \) (both 3.7- and 4.2-kb bands) were absent in the group \( D \) isolates, whereas a prominent band of \( \approx 4.4 \) kb was present in each (figure 1). The genotype \( D \) isolates were recovered from 5 diverse localities (Singapore, Australia, Japan, Iowa, and Israel), from HIV-infected patients (Australia and Iowa) and from patients not known to be infected with HIV or whose HIV status was unknown or unclear (Singapore, Ja-
pan, and Israel) (table 1). One type (D1) was found in 2 countries and the other (D2) in 4 countries.

Geographic and temporal differences. Analyses of the number of isolates within each of the major genotype groups from the various localities were undertaken to assess variations between regions and over time, as well as association with patients known to be infected with HIV. The numbers of isolates in each of the genotype groups A, B, and C were compared. The genotype D isolates were excluded from these analyses because of their rarity.

Comparison of all 15 localities revealed that significant differences existed in the distribution of isolates within genotype groups A, B, and C ($P < .0001$). To assess which localities differed, various combinations of localities were compared and combined. Analysis of the isolates examined in 2 early studies [9, 10] showed that these were indistinguishable ($P = .5$), and therefore the results of these “early studies” (of isolates obtained in 1985 or earlier, largely before 1982) were combined (table 1, lines 1 and 2). Analyses of the isolates from the 4 United States localities in the present study showed that these were indistinguishable ($P = 4$), and therefore these were combined as a group of overall “recent” isolates (obtained in 1994 or later) recovered in the United States (table 1, lines 12–15).

Comparison of the number of isolates in genotype groups A, B, and C from the United Kingdom (not analyzed in any previous study), Spain, France, Zimbabwe, Colombia, Singapore, Australia, and Japan revealed no statistically significant differences among these ($P = .2$). For comparison, the results for all these isolates were combined as a global group of “recent” isolates (obtained in 1985 or later, largely in 1995 or later) (table 1, lines 3–10).

Comparison of the distribution of isolates in genotype groups A, B, and C from the early studies [9, 10] with the overall recent United States isolates suggested a difference that did not reach statistical significance ($P = .057$). However, comparison of the

Table 1. Locality, number, and genotype distribution of global isolates of Candida albicans.

<table>
<thead>
<tr>
<th>Locale</th>
<th>Year</th>
<th>No. of isolates in genotype group</th>
<th>No. of DNA types (old/new) in genotype group</th>
<th>No. of DNA types unique to locale in genotype group</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA/UK</td>
<td>Pre-1982</td>
<td>91 78 13 0 0</td>
<td>20 0/9 0/0 0/0</td>
<td>9 1 0 0</td>
</tr>
<tr>
<td>Glasgow</td>
<td>1985</td>
<td>13 12 1 0 0</td>
<td>5 0/1 0/0 0/0</td>
<td>1 1 0 0</td>
</tr>
<tr>
<td>UK</td>
<td>1986</td>
<td>5 1 2 2 0</td>
<td>1 0/1 0/2 0/0</td>
<td>0 0 2 0</td>
</tr>
<tr>
<td>Spain</td>
<td>1985–90</td>
<td>54 32 15 7 0</td>
<td>7 1/3 3/7 0/3</td>
<td>8 5 0 0</td>
</tr>
<tr>
<td>France</td>
<td>1996</td>
<td>21 14 3 4 0</td>
<td>3 5/1 2/3 0/3</td>
<td>3 1 0 0</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>1996</td>
<td>19 12 6 1 0</td>
<td>2 6/3 2/0/1 0/0</td>
<td>4 2 1 0</td>
</tr>
<tr>
<td>Colombia</td>
<td>1996</td>
<td>20 16 2 2 0</td>
<td>6 2/1 1/2 0/2</td>
<td>1 1 2 0</td>
</tr>
<tr>
<td>Singapore</td>
<td>1995</td>
<td>23 8 7 7 1</td>
<td>3 2/3 3/0/3 0/1</td>
<td>2 2 1 0</td>
</tr>
<tr>
<td>Australia</td>
<td>1990–95</td>
<td>23 14 3 2 4</td>
<td>3 5/1 2/0 0/2</td>
<td>0 2 0 0</td>
</tr>
<tr>
<td>Japan</td>
<td>1996–97</td>
<td>52 30 9 12 1</td>
<td>8 6/5 3/7 0/1</td>
<td>4 2 4 0</td>
</tr>
<tr>
<td>Israel</td>
<td>1997</td>
<td>30 6 6 16 2</td>
<td>3 2/3 0/7 0/0</td>
<td>1 0 2 0</td>
</tr>
<tr>
<td>California</td>
<td>1994–97</td>
<td>17 14 3 0 0</td>
<td>5 2/1 1/0 0/0</td>
<td>0 1 0 0</td>
</tr>
<tr>
<td>Texas</td>
<td>1996</td>
<td>36 24 11 1 0</td>
<td>6 9/2 7/0 0/1</td>
<td>2 3 0 0</td>
</tr>
<tr>
<td>Iowa</td>
<td>1997</td>
<td>25 19 3 2 1</td>
<td>5 7/1 2/0 0/1</td>
<td>3 1 1 0</td>
</tr>
<tr>
<td>Georgia</td>
<td>1994</td>
<td>10 9 1 0 0</td>
<td>3 2/1 0/0 0/0</td>
<td>2 0 0 0</td>
</tr>
<tr>
<td>Totals</td>
<td>Pre-1990</td>
<td>163 123 31 9 0</td>
<td>27 6/5 4/7 0/1</td>
<td>27 16 54 9</td>
</tr>
<tr>
<td>Post-1990</td>
<td>276 166 54 47 9</td>
<td>60 6/5 4/7 0/1</td>
<td>60 47 54 9</td>
<td></td>
</tr>
<tr>
<td>Pre-1982 to 1997</td>
<td>439 289 85 56 9</td>
<td>40 6/5 4/7 0/1</td>
<td>40 34 54 9</td>
<td></td>
</tr>
</tbody>
</table>

* “New” refers to those types not represented in 33 originally detailed (9, 10, lines 1 and 2 of this table).
number of isolates in genotype groups A, B, and C from the recent global isolates showed that the recent global isolates were highly significantly different \((P < .0001)\) when compared with the isolates in early studies \([9, 10]\) and \(P = .0032\) when compared with the overall recent United States isolates.

Combination of the numbers of strains in genotype groups A, B, and C from the 2 early studies \([9, 10]\) and the overall recent United States isolates (from California, Texas, Iowa, and Georgia) yielded a total of 156 type A, 32 type B, and 3 type C isolates. Comparison of this combined United States/United Kingdom group with the recent global isolates showed that these 2 large groups were also significantly different \((P < .0001)\).

The distribution of isolates from Israel in genotype groups A, B, and C was significantly different from those of isolates in the early studies \([9, 10]\), the overall recent United States isolates, and the recent global isolates \((P < .0001\) for each analysis), as well as those from every other individual locality studied. In Israel, there was a high proportion of genotype group C isolates. This difference was not seen in the distribution between the distributions of isolates in genotype groups A, B, and C from Israel and those from the United Kingdom in 1986 \((P = .66)\). The small number of isolates from the United Kingdom \((n = 5)\) made this comparison unreliable.

The distribution of isolates in genotype groups A, B, and C isolated before 1990 \((n = 163)\) was compared with the distribution of those isolated since 1990 \((n = 276)\). This analysis revealed that these 2 sets were statistically different from each other \((P = .009)\). This temporal difference was due to the large increase in the isolation of genotype group C isolates \((5.5%\) before 1990 vs. \(17%\) after 1990).

The comparison of combined isolates known to come from patients infected with HIV \((n = 151)\) with those from patients who were known not to be infected with HIV or whose HIV status was unknown or unclear at the time of isolation of \(C.\ albidans\) \((n = 288)\) revealed no statistical difference between the numbers of isolates in genotype groups A, B, and C \((P = .995)\).

Expansion of the identified types of \(C.\ albidans\). The number of new types observed among the isolates from each locality was assessed. There was no statistically significant difference between the number of new genotypes observed for each locality and that originally observed in the 2 prior studies. The rate of new genotypes observed at each locality (i.e., the number of isolates that needed to be analyzed for each new type identified) varied only modestly between sites and was similar to the mean rate of new genotypes reported previously. Finally, statistically significant differences in the number of new genotypes per number of isolates were not observed if the isolates were divided temporally \((\text{pre-1990 vs. post-1990})\) or according to the HIV status of the subjects.

The overall rate (including the genotype D isolates) of a new genotype being observed in the present study was 1 new type for 3.81 isolates examined and, for group A, B, and C isolates, 1 new genotype for every 3.79 isolates typed. Since the rate of new genotypes in prior studies \([9, 10]\) was 1 per every 3.15 isolates studied, new genotypes are appearing slightly less frequently as more isolates are studied.

The reported incidence of 1 particular type \((A2)\) in an earlier study \([9]\) was very high \((37[40%]\) of 91). The high frequency of this particular type was not seen in the other localities: there were only 17 isolates of this type \((A2)\) found among the 348 isolates \((4.9%)\), a difference that was statistically significant \((P < .0001)\). The large number of isolates of this type in this previous study \([9]\) skews the number of isolates of this type among the pre-1990 isolates \((n = 40)\), as well as that among the non-HIV-associated isolates \((n = 52)\). As a result, the number of isolates of type A2 among the pre-1990 versus post-1990 isolates is significantly different \((P < .0001)\), as is that among known HIV-infected versus non-HIV–infected patients \((P < .0001)\).

Discussion

The results presented here reveal diversity in the genotype groups of \(C.\ albidans\) isolated from different geographic areas. Larger samples from each country would be more informative, and large geographic areas outside the United States remain unsampled. More clinical information about the patients from whom isolates are obtained is desirable, including site of disease and HIV status. Future enlargement of the database is desirable.

The most striking result was the differences observed between isolates from within the United States and those from other countries. The observed differences could be attributed to the number of genotype C isolates outside the United States, whereas only 3 group C isolates were found within the United States, all of which had been isolated recently.

\(C.\ albidans\) isolated from global locations, including Africa, South America, Europe, Asia, and Australia, showed a similar genotypic distribution. However, this distribution was distinct from that found for the United States and Israel.

The observed differences for the United States, Israel, and other localities, as well as the temporal variation observed \((\text{pre-1990 vs. post-1990})\), could all be attributed to the changes in the number of isolates of genotype C. It would appear that this newly described \(C.\ albidans\) genotype \([11, 12]\) is considerably widespread and occurs more frequently among recent isolates. Genotype C may be increasing in occurrence globally, or, alternatively, this genotype now may be appearing only within the United States. This is the first report concerning use of this methodology on isolates from many of these countries. Thus, it is unknown whether genotype C isolates existed in these countries some time ago or have recently been acquired.

It was not possible in the present study to show that this change is attributable to the advent of HIV infection (and as-
associated changes in antifungal use), because no statistical difference was observed in the analysis of patients known to be infected with HIV. However, the analysis of these data could have been affected by the uncertainty of the HIV status of many of the patients included in this study.

It is possible that the ecology within the United States is changing, in some undefined way, so that group C isolates, which are more prevalent outside the United States, are now penetrating the United States pool from abroad. The nature of group C isolates, whether a hybrid of groups A and B or a transition form between them, remains open for speculation. Genotype C strains were shown [13] to have a transposable intron incompletely inserted into the rDNA repeats; this results, on PCR of the 25S rDNA, in a single ampliter from group A (450 bp) and B (850 bp) isolates and 2 ampliters (450 and 850 bp) from the group C isolates [13]. A number of different mechanisms for genomic variation in C. albicans, such as ploidy fluctuations, translocation, mitotic recombination, and non-disjunction, have been identified [19]. Recent studies [20] have suggested that, although populations of C. albicans are primarily clonal, evidence for recombination suggests that sexual reproduction or some form of genetic exchange occurs in this species.

All the isolates included in this study were defined to the species level as C. albicans by routine clinical laboratory methods. An assumption at the start of these studies was that the new genotype groups observed (C and D) are not representative of different species. However, this could have been an incorrect assumption, since the newly described species Candida dubliniensis is largely phenotypically indistinguishable from C. albicans [5, 21]. Furthermore, Candida stellatoidea, thought to be synonymous with C. albicans [22, 23], could have been represented by these genotypes. Evidence of this is found in an early description of the rDNA pattern of Candida stellatoidea [7] that appears to be the same or similar to that described for our C. albicans genotype B [8].

Concurrent research was undertaken [13] to establish the association of these genotypes (C and D) with non-albicans species of Candida. This concurrent research showed that C. stellatoidea is synonymous with C. albicans genotype B and that “C. albicans genotype D” is in fact C. dubliniensis [13]. Thus, this study also indicates geographic areas from which C. dubliniensis has not previously been reported. We also suggest that there are subtypes of C. dubliniensis, just as was shown with C. albicans. If this observation is expanded further, it may prove useful in future epidemiological studies of C. dubliniensis.

A correlation has been made between the genotype B isolates and an increased susceptibility to the antifungal agent flucytosine [9, 24]. Concurrent research [13] corroborated this association in the geographically diverse strains studied here and suggests that the newly described genotype C, as well as genotype D (C. dubliniensis), might be associated with changes in antifungal susceptibility. Selective pressures due to drug exposure could be a factor in ecological and temporal change. It may well be that the geographic genotypic differences observed represent some nonclonal evolution in response to environmental pressures such as antifungal drug use, which differs in different parts of the world, or changes in the ecology, such as the effect of rapid expansion of susceptible populations (immunocompromised hosts). Geographic genotypic differences could be used as a factor in local choices of antifungal drugs for prophylaxis or therapy. More information, other than drug susceptibility, concerning the characteristics of the 3 lineages of C. albicans may be gleaned by techniques that provide data on genome sequences. An association of different genotype groups with body sites has previously been indicated [9].

Acknowledgments

We thank Drs. Jeannette Y. Lee (University of Alabama, Birmingham) and Lynda Treat-Clemens (diaDeXus, LLC, Santa Clara, CA) for advice on the statistical analysis.

References

13. McCullough MJ, Clemens KV, Stevens DA. Molecular and phenotypic characterization of genotypic Candida albicans subgroups and comparison...