JOURNAL OF CLINICAL MICROBIOLOGY, Aug. 2002, p. 2729–2740 Vol. 40, No. 8
0095-1137/02/$04.00+0 DOI: 10.1128/JCM.40.8.2729–2740.2002

Ca3 Fingerprinting of *Candida albicans* Bloodstream Isolates from the United States, Canada, South America, and Europe Reveals a European Clade

Claude Pujol,[1] Michael Pfaller,[2] and David R. Soll[1]*

Departments of Biological Sciences[1] and Pathology,[2] University of Iowa, Iowa City, Iowa 52242

Received 18 March 2002/Returned for modification 6 May 2002/Accepted 13 May 2002

It was previously demonstrated by a cluster analysis that 26 unrelated U.S. isolates of *Candida albicans* separated into three distinct groups (groups I, II, and III) while South African isolates separated into four distinct groups (groups I, II, III, and SA). To verify the absence or underrepresentation of SA isolates in North America, and to identify which groups are represented in Europe and South America, collections of bloodstream isolates from each geographical locale were analyzed by cluster analyses based on genetic fingerprinting with the Ca3 probe. The results verify that North America is almost devoid of SA isolates (2%). However, the results reveal a new clade, designated group E, relatively specific to Europe. While 26% of a European collection of 46 isolates was composed of group E isolates, only 2% of the 164 North American isolates, 5% of 22 South American isolates, and 1% of 361 South African isolates were composed of group E isolates. The North American collection proved to be the least-diverse collection in regard to group representation. In a comparison of collections from the Northeast, Midwest, and Southwest regions of the United States, Canada, and South America, it was demonstrated that both the U.S. Southwest and the South American collections were devoid of group II isolates. Together these results identify for the first time a European-specific clade and demonstrate clear distinctions in the representations of the five demonstrated clades (groups I, II, III, SA, and E) in different geographical locales.

Using Southern blot hybridization with the DNA fingerprinting probe Ca3, we previously demonstrated that a random collection of 26 unrelated *Candida albicans* isolates from the United States separated into three discernible groups, with 35% of the isolates in group I, 31% in group II, and 19% in group III (9). The remaining 15% of U.S. isolates fell into no discernible group (outliers). The validity of the three groups was tested by performing a similar cluster analysis on the same collection of isolates with two unrelated DNA fingerprinting methods, the random amplified polymorphic DNA method and multilocus enzyme electrophoresis (9). These additional methods separated the isolates into approximately the same three groups and identified as outliers the same isolates as did the Ca3 fingerprinting method, in essence cross-verifying the effectiveness of all three methods for cluster analyses. Subsequently, Lott et al. (7) used an additional method to verify the three major groups, I, II, and III.

To test whether *C. albicans* from other geographical regions separated into the same three major groups, Blignaut et al. (3) used Southern blot hybridization with the Ca3 probe in a cluster analysis of 361 isolates collected from human immunodeficiency virus (HIV)-positive black South Africans and control groups of healthy white and black South Africans. In this study (3), mixed dendrograms of the South African isolates and the 26 U.S. isolates used in the original study by Pujol et al. (9) were generated to identify South African isolates in groups I, II, and III. The results of the study by Blignaut et al. (3) demonstrated that South African isolates did separate into groups I, II, and III. In addition, 53% of isolates from HIV-positive and healthy black South Africans and 33% of those from healthy white South Africans separated into a fourth clade not represented in the collection of 26 U.S. isolates. This fourth South African-specific clade was named group SA (3). Here we used Southern blot hybridization with the Ca3 probe to perform cluster analyses on 164 bloodstream isolates from the United States and Canada in order to corroborate the observation that group SA isolates are absent from North America. In addition, we performed cluster analyses on 46 isolates from Europe and 22 isolates from South America to test whether the SA clade is present in these geographical locales. Our results first confirm the finding by Blignaut et al. (3) that the SA clade is underrepresented in North America (2%) and, in addition, is underrepresented in South America (2%). The SA clade is, however, represented in Europe (10%). In addition, our results demonstrate for the first time the existence of a European clade (group E) that makes up approximately 26% of European isolates but only 2% of North American isolates, 5% of South American isolates, and 1% of South African isolates.

MATERIALS AND METHODS

Collection and maintenance of isolates. All new isolates in this study were collected from the bloodstreams of patients in 42 different geographical sites participating in the SENTRY Surveillance Program (8). Isolates were initially plated on blood agar and Sabouraud dextrose agar at the original site of collection and then sent to the University of Iowa Hospitals and Clinics for banking and further analysis. Upon receipt, each isolate was subcultured onto potato dextrose agar (Remel, Lenexa, Kans.) and CHROMagar (Hardy Diagnostics, Santa Maria, Calif.) to assess viability and species homogeneity. *C. albicans* isolates were identified with Vitek and API kits (bioMerieux, St. Louis, Mo.) and

* Corresponding author. Mailing address: Department of Biological Sciences, University of Iowa, Iowa City, IA 52242. Phone: (319) 335-1117. Fax: (319) 335-2772. E-mail: david-soll@uiowa.edu.

TABLE 1. Collection sites and number of isolates fingerprinted

Collection name	Site no.	City (state, province, or country)	Region[a]	No. of isolates fingerprinted
North America (United States and Canada)	1	Wilmington (Del.)	EC	6
	2	Indianapolis (Ind.)	MW	4
	4	Akron (Ohio)	MW	2
	8	Albuquerque (N.M.)	SW	6
	11	Iowa City (Iowa)	MW	18
	12	Omaha (Nebr.)	MW	3
	14	Boston (Mass.)	EC	6
	17	Rochester (N.Y.)	EC	4
	19	Berkeley (Calif.)	WC	2
	21	Seattle (Wash.)	WC	1
	23	Dallas (Tex.)	SW	2
	24	Houston (Tex.)	SW	5
	25	Galveston (Tex.)	SW	22
	27	Louisville (Ky.)	MW	2
	30	Charlottesville (Va.)	EC	15
	51	Salt Lake City (Utah)	MW	3
	52	Burlington (Mass.)	EC	3
	55	Miami Beach (Fla.)	EC	4
	82	New York (N.Y.)	EC	31
	31	Edmonton (Alberta)	Can	2
	32	Winnipeg (Manitoba)	Can	6
	33	Halifax (Nova Scotia)	Can	4
	35	Ottawa (Ontario)	Can	7
	38	Montreal (Quebec)	Can	6
Europe	58	Brussels (Belgium)		1
	61	Lille (France)		7
	62	Athens (Greece)		1
	64	Seville (Spain)		3
	66	Madrid (Spain)		5
	67	Lausanne (Switzerland)		7
	84	Genoa (Italy)		6
	86	Rome (Italy)		1
	88	Frankfurt (Germany)		8
	89	Linkoping (Sweden)		2
	91	London (United Kingdom)		5
South America	44	Medellin (Colombia)		1
	48	São Paulo (Brazil)		12
	49	Caracas (Venezuela)		4
	57	Porto Alegre (Brazil)		5
Turkey and Israel	68	Ankara (Turkey)		3
	69	Istanbul (Turkey)		3
	63	Tel-Hashomer (Israel)		5

[a] EC, East Coast; MW, Midwest; SW, Southwest; WC, West Coast; Can, Canada.

other conventional methods as required (16). Clonal isolates were stored as water suspensions at ambient temperature.

Isolates were collected from 19 sites in the United States and 5 sites in Canada (Table 1). Together these isolates made up the North American collection and included 164 isolates. When highly related isolates from the same hospital (i.e., when the similarity coefficient between the patterns of a pair of isolates A and B [S_{AB}], further defined below, was >0.90) were considered the same strain and counted only once, the North American collection was reduced to 122 isolates. Isolates collected from 11 sites in Europe totaled 46 and were reduced to 37 when highly related isolates from the same hospital were counted only once (Table 1). Isolates collected from four sites in South America totaled 22 and were reduced to 19 when highly related isolates from the same hospital were counted only once (Table 1). Finally, isolates collected from three sites in Turkey and Israel totaled 11.

DNA fingerprinting. All isolates were fingerprinted by Southern blot hybridization using the complex DNA fingerprinting probe Ca3 (1, 6, 10, 12) according to methods previously described in detail (14, 15). In brief, DNA was extracted from cells (13) and digested with *Eco*RI. DNA was then electrophoresed in a 0.8% agarose gel, transferred to Hybond N^+ membrane (Amersham, Piscataway, N.J.) by blotting, prehybridized with salmon sperm DNA, hybridized overnight with a ^{32}P-labeled Ca3 probe, and autoradiographed.

Computer-assisted cluster analysis. Autoradiogram images were digitized into the DENDRON software database (15). By use of this software program, vertical, horizontal, and nonlinear distortions were removed with the unwarping options. Processed hybridization patterns were then automatically scanned to identify all bands and to link common bands (15). The patterns of all test isolates were then compared in a pairwise fashion, and the S_{AB}s were computed according to the formula $S_{AB} = 2E/(2E + a + b)$, where E is the number of bands common to strains A and B, a is the number of bands unique to strain A, and b is the number of bands unique to strain B. An S_{AB} of 0.00 represents total unrelatedness between isolates A and B, an S_{AB} of 1.00 represents the identical match of all bands between isolates A and B, and increasing values of S_{AB} from 0.01 to 0.99 represent increasing levels of similarity. An S_{AB} threshold of 0.7 was selected for determining groups in cluster analyses. This value, which has proved effective in separating groups in previous cluster analyses employing Ca3 fingerprinting methods (2, 7), is just above the average S_{AB} of the tested collections. Band data from isolates analyzed in previous studies (3, 9) had been stored in the DENDRON database and were retrieved to generate mixed dendrograms with new data according to methods described earlier (15).

IS*1* sequence analysis. The IS*1* sequences of test isolates were amplified by the PCR method (Labline, Melrose Park, Ill.) according to the methods of Lott et al. (7). The IS*1* primers used in amplification were 5′-GGGAATCTGACTGTCT AATTAA-3′ and 5′-CTTGGCTGTGGTTTCGCTAGAT-3′. Following initial incubation at 95°C, 35 cycles of 1-min steps were run at 95, 55, and 72°C. Final elongation was performed for 5 min at 72°C.

RESULTS

Method for identifying groups. The set of 26 isolates characterized by Pujol et al. (9) was used as a reference collection to identify members of groups I, II, and III in other collections by the following protocol. Isolates in the new collection were DNA fingerprinted with the complex DNA fingerprinting probe Ca3, and the pattern data for each strain was stored in the DENDRON database (15). The DNA fingerprinting data of the reference collection, which is also stored in the DENDRON database, were then combined with data from the new collection, and S_{AB}s were computed for all pairs of isolates in the combined collection. A mixed dendrogram was then generated between reference and new isolates based on S_{AB} values. Because the isolates in the reference collection had already been grouped, the isolates in the new collection could readily be identified in groups I, II, and III through coclustering characteristics. In this case, groups were readily identified at an S_{AB} threshold of 0.70, a value close to the average for unrelated isolates. By use of this method, the SA group was identified in South African collections by Blignaut et al. (3) (Fig. 1B). Here we first expanded the reference collection to include not only the 26 U.S. reference isolates (9) (Fig. 1A) but also SA isolates from healthy black individuals attending the Medunsa Clinic in South Africa (3) (Fig. 1B). A dendrogram of this expanded reference collection, with the four groups (I, II, III, and SA) delineated, is presented in Fig. 1C.

The European collection contains a European-specific clade (group E). The European collection included 46 bloodstream isolates collected from 11 sites spanning nine countries (Table 1). A mixed dendrogram based on S_{AB}s computed between all pairwise combinations of European and reference isolates identified isolates in groups I, II, III, and SA (Fig. 2A). The proportions in the four groups were 20, 9, 7, and 11%, respectively (Table 2). In addition, a fifth major group that included 12 European isolates (26%) emerged in the collection. No outliers in the reference collection coclustered with isolates in this new cluster (Fig. 2A). We refer to this new clade as group E. A dendrogram generated in the absence of reference iso-

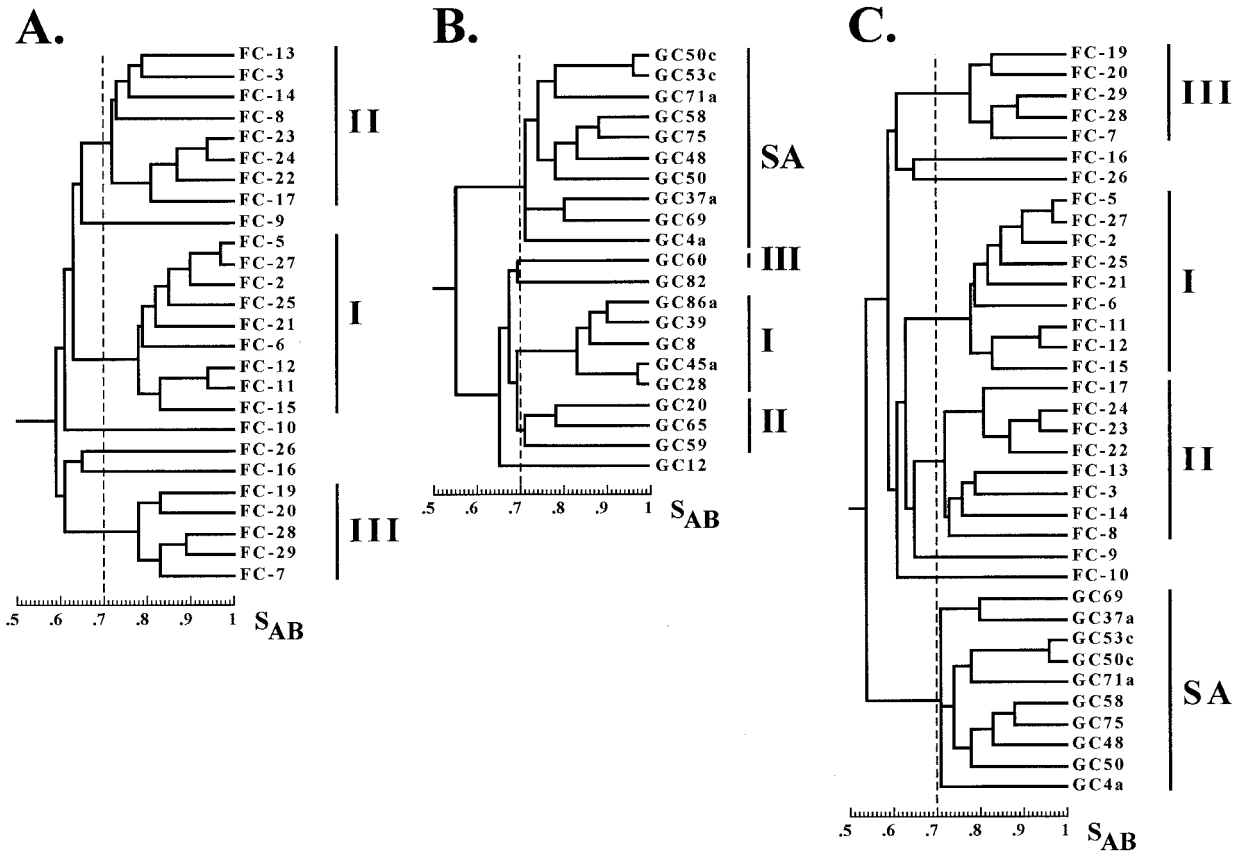

FIG. 1. Reference isolates from previous studies used in mixed dendrograms to identify isolates in new collections that are members of groups I, II, III, and SA. (A) A dendrogram of the collection of 26 isolates from the United States originally characterized by Pujol et al. (9) in the identification of groups I, II, and III. (B) A dendrogram of the collection of 21 isolates from healthy black South Africans attending the Medunsa Clinic in South Africa characterized by Blignaut et al. (3), which contains group SA isolates in addition to group I, II, and III isolates. (C) A mixed dendrogram of the collection of 26 U.S. isolates from panel A and only the SA isolates from panel B, which represents the reference collection for generating mixed dendrograms with new collections in order to identify isolates from groups I, II, III, and SA.

lates revealed three or more European isolates in groups I, II, III, SA, E, and three additional groups, e1, e2, and e3, defined by the S_{AB} threshold of 0.70, (Fig. 2B). When isolates from the same hospitals with S_{AB}s >0.90 were considered a single isolate and a dendrogram was generated from the reduced collection, isolates separated into groups I, II, III, SA, E, and the three additional European subgroups (Fig. 2C). It should be noted that Groups e1, e2, and e3 were most closely related to group III (Fig. 2C). Group E isolates made up 22% of the reduced European collection (Table 2), whereas isolates from groups e1, e2, and e3 each made up 7% of the collection.

The North American collection contains few SA or E isolates. To test whether group SA isolates could be identified in an expanded U.S. collection and whether E isolates were represented, we analyzed a collection of 139 U.S. and 25 Canadian bloodstream isolates considered together to represent the North American collection (Table 1). To identify isolates from groups I, II, III, SA, and E in the U.S. collection, a mixed dendrogram was first generated based on the S_{AB}s between the combined isolates in the North American collection, the reference collection of U.S. isolates, group SA isolates from the Medunsa Clinic in South Africa (Fig. 1C), and the 12 group E isolates from the European collection (Fig. 2B). Because of its length, this mixed dendrogram has been separated into the top (Fig. 3A), middle (Fig. 3B), and bottom (Fig. 3C) portions to accommodate it on a single page of text. Seventy-seven isolates (47%) of the North American collection separated into group I, 20 (13%) into group II, and 39 (24%) into group III. Only four isolates (2%) separated into group SA, and four isolates (2%) separated into group E. Twenty-four isolates did not fall into the four major groups (Fig. 3). When the North American collection was reduced by counting isolates from the same hospitals with S_{AB}s >0.90 as one isolate, the proportions of the five groups were similar (Fig. 4; Table 2). In the reduced North American collection of 122 isolates, 20 (16%) did not separate into the four major groups (Fig. 4; Table 2). Eight of these outliers separated into two clusters of three or more isolates, three in subgroup na1 and five in subgroup na2. To test whether these two minor groups were related to the three minor European groups e1, e2, and e3, or for that matter, whether any European isolates were related to isolates in these North American subgroups, we generated a mixed dendrogram between the reduced European and reduced North American collections. Two North American outliers (P57049 and P76001) that did not reside in either na1 or na2 coclustered with group e1 isolates, and one European outlier (P22089) that did not reside in e1, e2, or e3 coclustered with na2 isolates (data not shown). Using Fisher's exact test, signif-

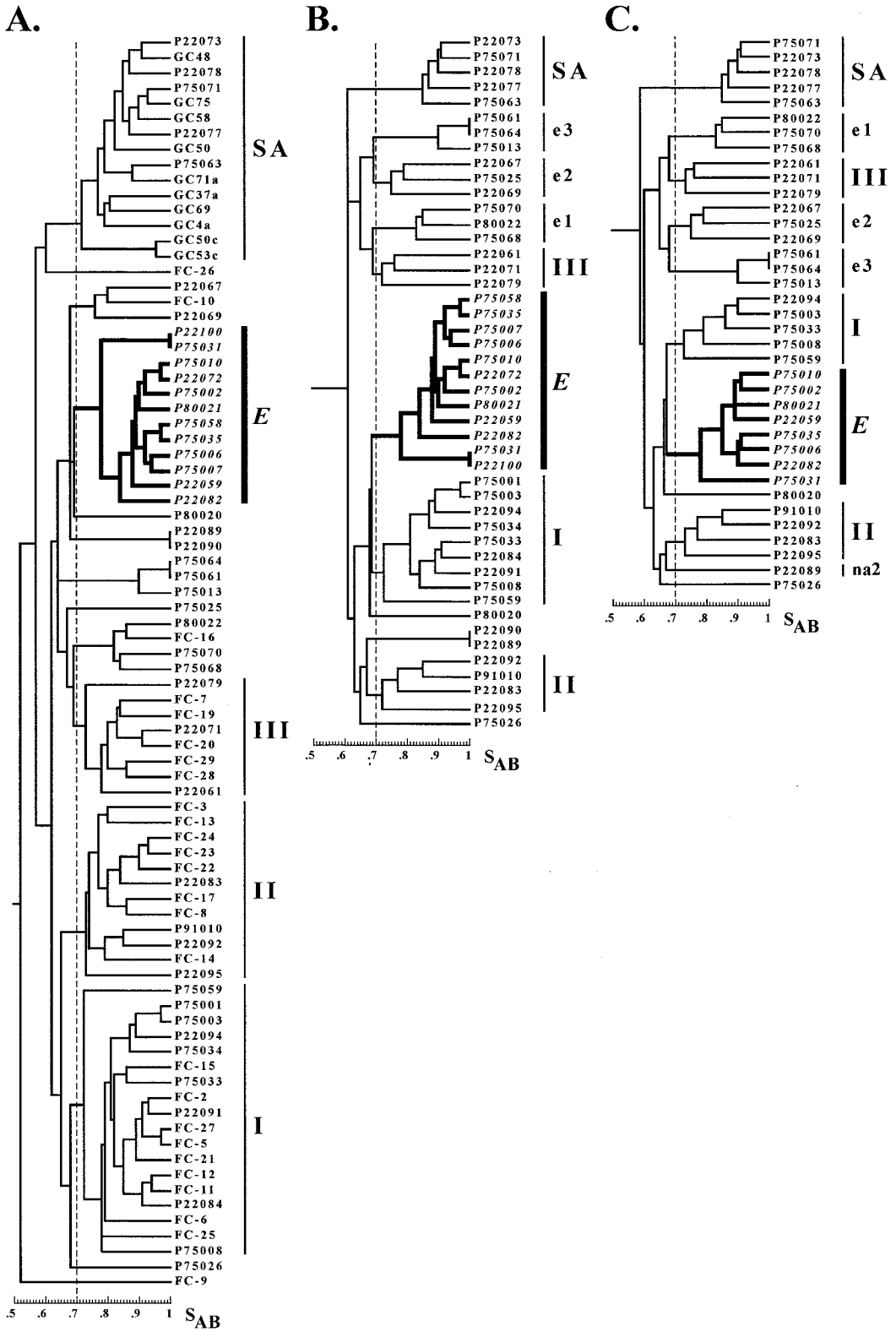

FIG. 2. The European collection contains a new clade, group E. (A) A mixed dendrogram generated with the 46 isolates in the European collection (Table 1) and the reference collection illustrated in Fig. 1C, in which groups I, II, III, SA, and the new group E are represented. (B) The European collection alone. (C) The reduced European collection, in which isolates from the same hospital with S_{AB}s >0.90 are considered a single isolate. Reference isolates are boxed in gray. Isolates are demarcated in minor European groups e1, e2, and e3 and the minor North American group na2.

TABLE 2. Proportions of *C. albicans* collections in the different groups defined in cluster analyses based on DNA fingerprinting with the Ca3 probe

Isolates included	Geographical region	No. of isolates	% of isolates classified as:					
			Group I	Group II	Group III	Group SA	Group E	Outliers
All isolates[a]	United States, East Coast	69	53	13	17	3	3	13
	United States, Midwest	32	47	19	13	0	3	19
	United States, Southwest	35	37	0	51	0	3	9
	Canada	25	40	29	24	8	0	4
	North America (United States and Canada)	164	47	13	24	2	2	11
	Europe	46	20	9	7	11	26	9
	South America	22	54	0	23	0	5	18
	Turkey - Israel	11	18	0	0	45	0	36
Isolates with threshold S_{AB}s of <0.90[b]	North America (reduced)	122	38	14	27	2	3	15
	Europe (reduced)	37	13	11	8	13	22	32
	South America (reduced)	19	53	0	21	0	5	21

[a] The analysis included some isolates from the same hospitals with very high S_{AB}s that could represent the same endemic strain.
[b] The analysis included only isolates with S_{AB}s less than 0.90, the threshold for highly related isolates under the conditions employed for fingerprinting (15).

icant differences were demonstrated in the proportions of isolates in the different groups. This was true for the general distribution of isolates ($P < 10^{-3}$) or when groups were considered separately (group I, $P = 1.3 \times 10^{-2}$; group E, $P = 2 \times 10^{-6}$). This was also true for the reduced collections.

Differences between geographical regions in North America. The collection of isolates in North America was large enough to be separated into the following four regions: East Coast ($n = 69$), Midwest ($n = 32$), Southwest ($n = 35$), and Canada ($n = 25$) (Tables 1 and 2). Three West Coast isolates (Table 1) were not included in this analysis. The collections from the East Coast, Midwest, and Canada subdivided at roughly the same proportions into groups I, II, III, SA, and E (Table 2). Using Fisher's exact test, the distribution of isolates into the different groups was not found to be significantly different among the three populations ($P > 0.05$). This was also true when the reduced collections were used. The Southwest collection, however, differed (Southwest and East Coast, $P = 3 \times 10^{-3}$; Southwest and Midwest, $P = 10^{-3}$; Southwest and Canada, $P = 3 \times 10^{-3}$). It contained no group II isolates and an extremely high proportion of group III isolates (Table 2). These differences were found to be significant when the Southwest collection was compared to the three other North American collections individually or when the latter were grouped together (P values ranged from 3×10^{-2} to 8×10^{-5}). The majority of the differences remained significant when reduced collections were considered. In only one case, Southwest versus East Coast, did the difference lose significance, probably reflecting the fact that group II isolates were less prevalent in the East Coast than in Canada and the Midwest.

The South American collection. The South American collection was primarily from Brazil, with limited isolates from Venezuela and Colombia. Of the reduced collection of 19 isolates, 53% were in group I, 0% in group II, 21% in group III, 0% in group SA, and 5% in group E, while 21% were outliers (Table 2). The most noteworthy characteristics of the collection were the absence of group II isolates, which was also observed in the U.S. Southwest, and the negligible levels of group SA and group E isolates, both characteristics of the general North American collection. When entire collections were compared, the South American one did not differ significantly from those of the Southwest, Midwest, or East Coast. It did differ significantly from that of Canada ($P = 3.2 \times 10^{-2}$). This difference was due to the lack of group II isolates in the South American collection. When the proportions of group II isolates were compared, significant differences were observed between South America and Canada or the Midwest ($P < 0.05$), but not between South America and the Southwest or East coast, again suggesting that group II isolates are more prevalent in Canada and the Midwest. The South American collection also differed significantly from the European collection ($P = 3 \times 10^{3}$).

The Turkey and Israel collection. Although the Turkey and Israel collection included only 11 isolates, it exhibited one noticeable characteristic, a high proportion of group SA isolates (45%). This proportion was within the range observed in South Africa for blacks (55%) and whites (33%) (3) but higher than that in the European collection (11%) ($P = 1.5 \times 10^{-2}$) and far higher than that in the North American collection (2%) ($P = 3.8 \times 10^{-5}$) and South American collection (0%) ($P = 1.9 \times 10^{-3}$) (Table 2).

The South African collection is almost devoid of group E isolates. Blignaut et al. (3) demonstrated that isolates from HIV-positive black South Africans separated into group I (18%), group II (17%), group III (4%), and group SA (55%), while 7% were outliers. Similar proportions (21, 11, 8, 53, and 8%, respectively) were obtained for isolates from healthy black South Africans (3). Isolates from healthy white South Africans separated into the same four major groups, but the proportions differed from that of the black South African populations. A greater proportion of isolates from white South Africans clustered in group III (20 versus 6%), while a smaller proportion clustered in group SA (33 versus 53%). In the white South African collection, 13% were outliers. To test whether any of the South African outliers clustered in group E, mixed dendrograms were generated: first, of the collection of 21 isolates from black control individuals attending the Medunsa clinic and the 12 group E isolates from the European collection (Fig. 5A), and second, of the collection of 46 isolates from white control individuals from Pretoria and the 12 group E isolates

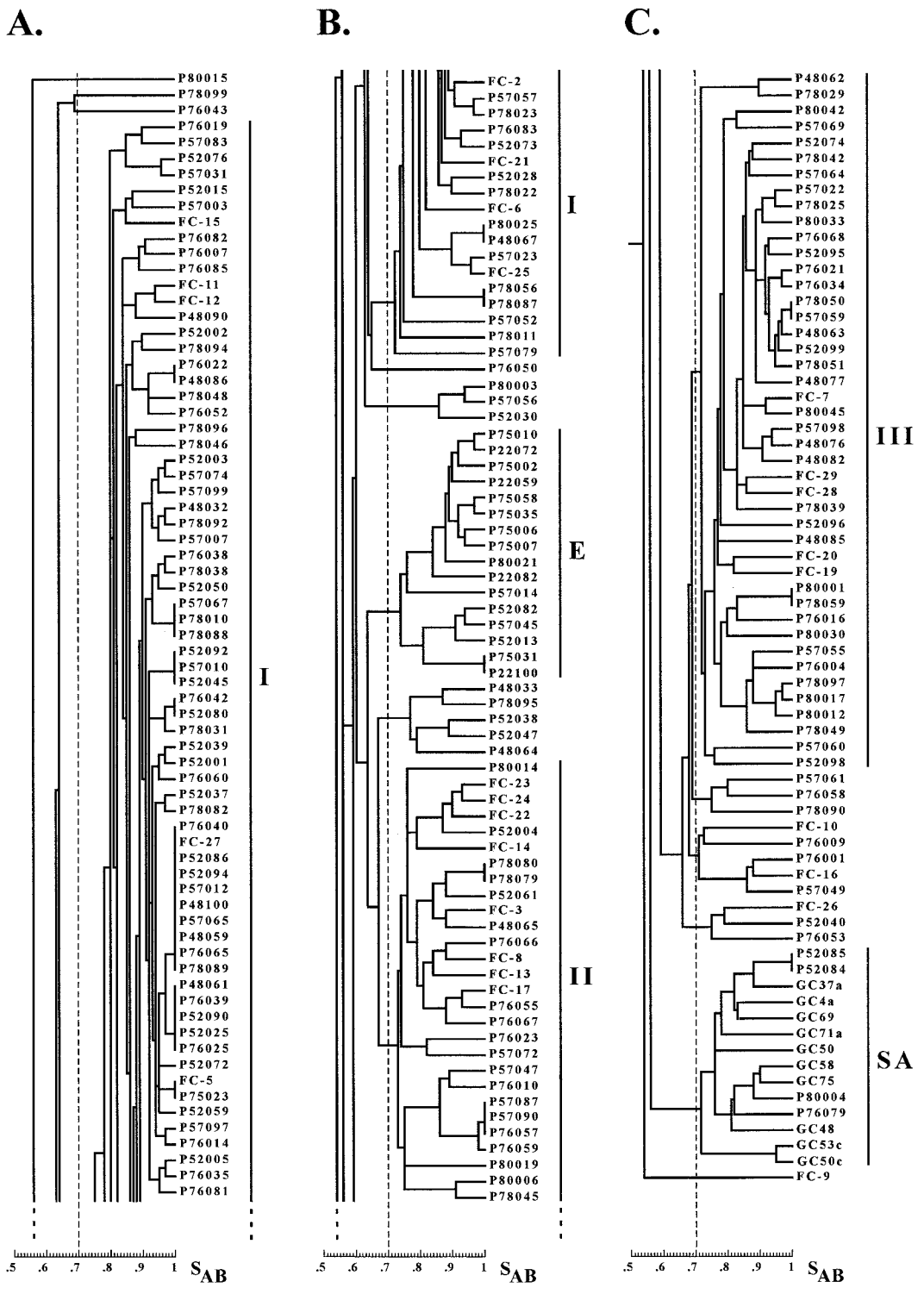

FIG. 3. The total North American collection. The mixed dendrogram was generated with the 164 isolates in the North American collection (Table 1), the reference collection in Fig. 1C, and the 12 group E isolates from the European collection. The dendrogram, which includes 212 isolates, has been separated into top (A), middle (B), and bottom (C) portions to fit on one page. Reference isolates are boxed in gray.

from the European collection (Fig. 5B). None of the black Medunsa collection isolates coclustered with group E isolates (Fig. 5A), and only one of the white Pretoria collection (2%) coclustered with group E isolates (Fig. 5B). When a mixed dendrogram was generated with all 361 South African isolates of the Blignaut et al. (3) study and the 12 group E isolates from the European collection, only five of the South African isolates (1%) coclustered with the E isolates (data not shown). These

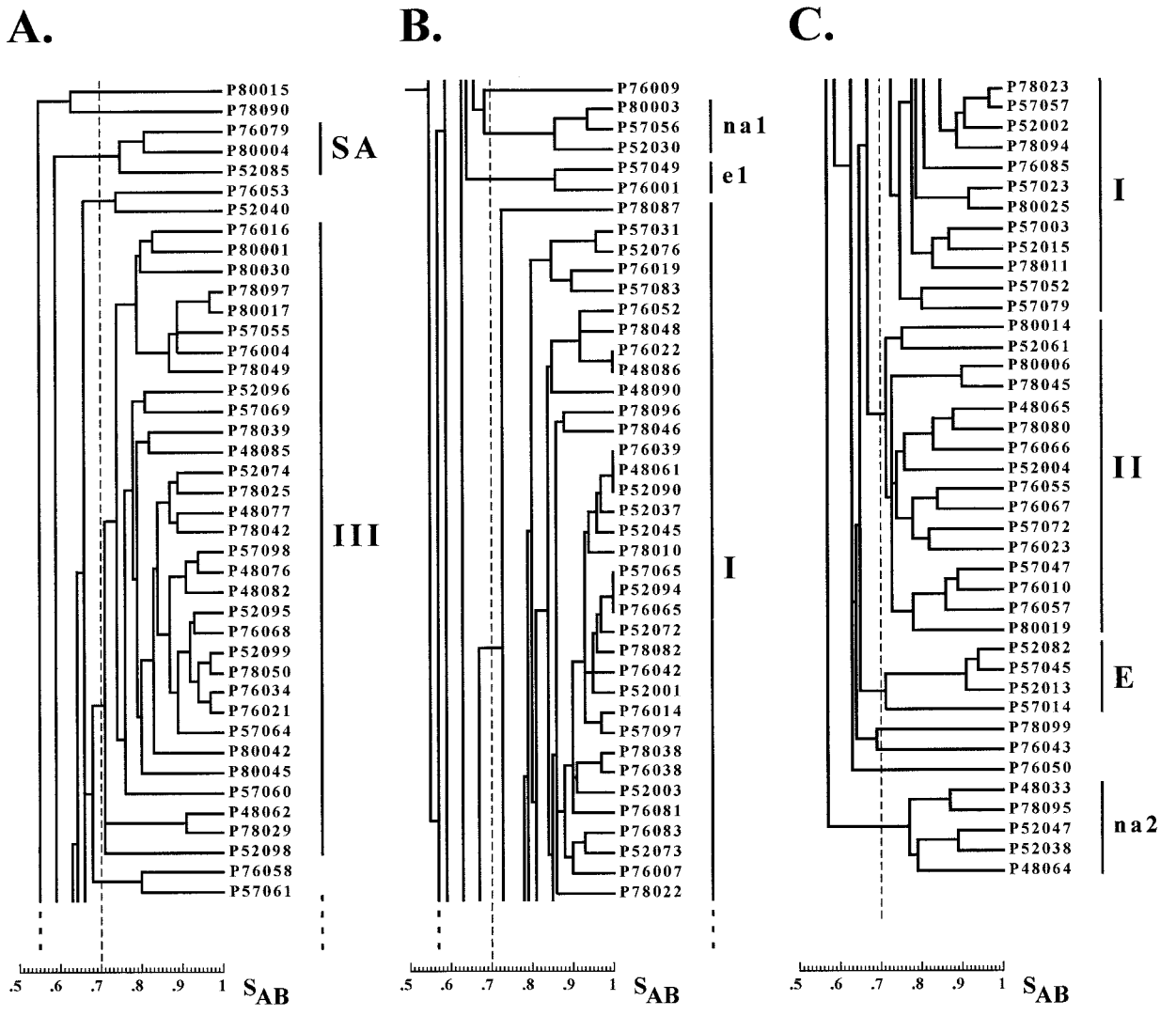

FIG. 4. The reduced North American collection. A dendrogram of the reduced North American collection, in which isolates from the same hospital with $S_{AB}s > 0.90$ are considered a single isolate, was generated. The dendrogram, which includes 122 isolates, has been separated into top (A), middle (B), and bottom (C) portions to fit on one page. In addition to groups I, II, III, SA, and E, minor groups na1, na2, and e1 are demarcated.

results demonstrate that group E isolates are dramatically underrepresented not only in North America and South America but also in South Africa.

Group E fingerprint patterns. It was previously demonstrated that the node separating the group SA cluster from groups I, II, and III was the most deeply rooted in dendrograms of South African isolates alone or mixed with U.S. isolates (3). In contrast, the node separating group E isolates from other groups in the dendrogram for European isolates was never as deep as those distinguishing groups SA and III and was roughly in the range of the node separating groups I and II (Fig. 2). When modeled DNA fingerprint patterns (15) of group E isolates were compared with those of group I, II, III, and SA isolates, distinguishing features were identified. Like group I, II, III, and SA patterns, group E patterns included conserved bands at 2.7, 3.3, 3.8, 4.5, 5.4, and 19.0 kb (Fig. 6). However, group E patterns exhibited a decrease in the number of bands in the range of 5.4 to 19.0 kb, an increase in the number of bands in the range of 3.8 to 4.5 kb range, and the presence in all tested strains of bands at 2.0 and 2.4 kb (Fig. 6).

Group E isolates contain the IS*1* intron. The presence of the IS*1* intron in the 25S rRNA gene has been demonstrated to be a characteristic of both group III (7) and group SA (3) isolates. Its absence in groups I and II has been demonstrated (7). To test whether group E isolates contain the IS*1* intron, we tested 21 group E isolates from around the world using PCR with customized primers (7). The size of the amplification product containing the 379-bp IS*1* group I intron was 626 bp, and without the intron it was 247 bp. In Fig. 7A, a representative gel of the PCR products for group I, II, III, SA, and E isolates is presented, and in Fig. 7B a dendrogram and synopsized data for the 21 tested group E isolates are presented. Nine of the 21 group E isolates contained the IS*1* intron (Fig. 7B). In Fig. 7A, examples of group E isolates with and without the IS*1* intron are presented. In the dendrogram of the group E isolates in Fig. 7B, two subgroups emerged, subgroup a, which contained

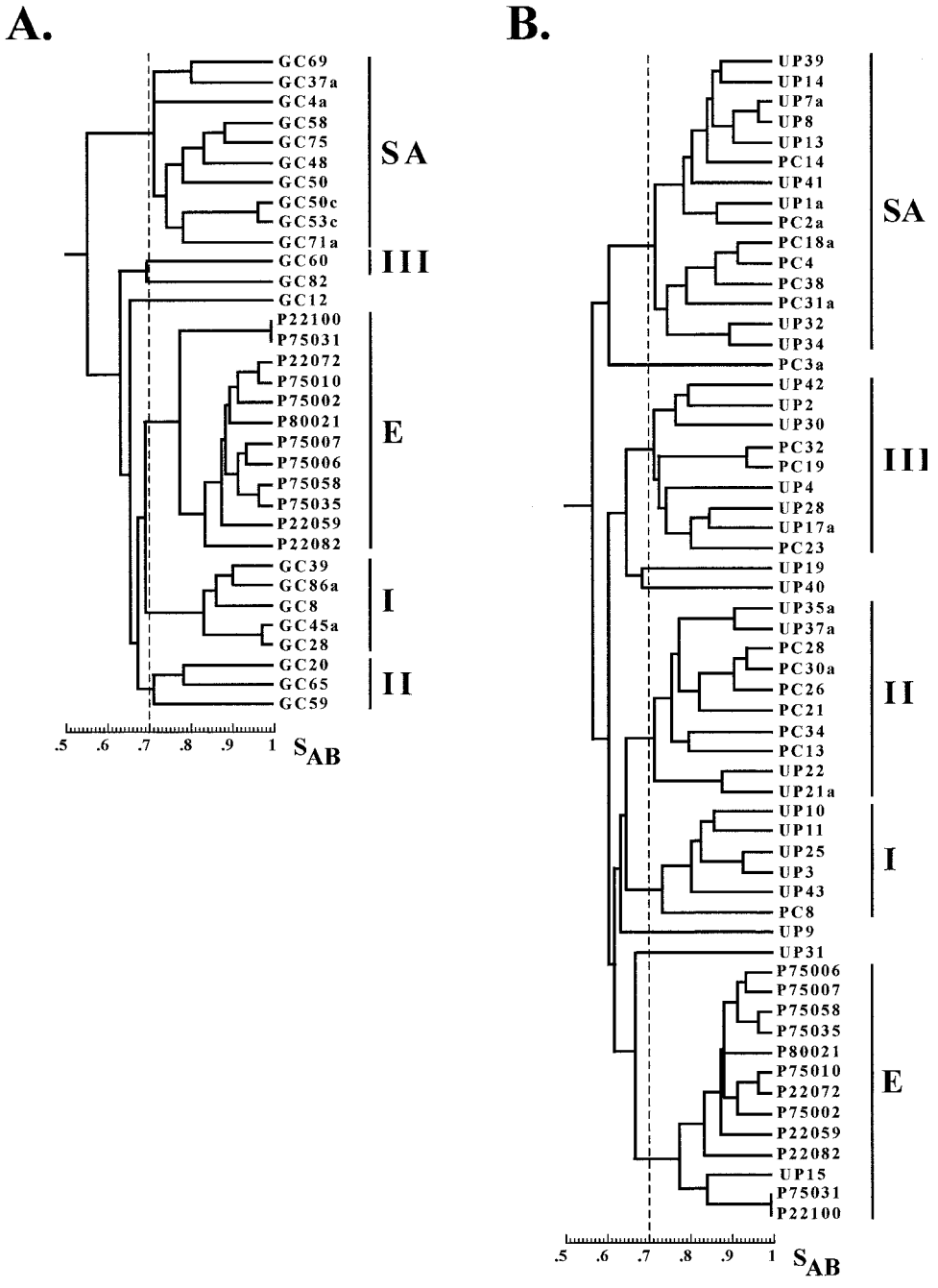

FIG. 5. South Africa is relatively devoid of group E isolates. (A) A mixed dendrogram generated with 21 isolates from healthy black South Africans attending the Medunsa Clinic and the 12 group E isolates from the European collection. (B) A mixed dendrogram generated with 46 isolates from healthy white South Africans in Pretoria and the 12 group E isolates from the European collection. European reference isolates are boxed in gray. Note that no Medunsa isolate and only one from Pretoria coclusters with European group E isolates.

the majority of tested European isolates, and subgroup b, which was more heterogeneous and contained roughly equal mixes of South African, U.S., and European isolates (Fig. 7B). The IS*1* intron was restricted to subgroup b isolates (Fig. 7B).

DISCUSSION

Because the mode of reproduction of *C. albicans* is essentially clonal (2, 11), isolates can be separated into genetic groups. Attempts to accomplish this began prior to the availability of DNA fingerprinting technologies. More than 40 years ago, Hasenclever and Mitchell (4, 5) separated *C. albicans* into serotypes A and B, based on differences in polysaccharide moieties on the cell surface. However, even with the development of effective genetic fingerprinting methods in the late 1980s, genetically related groups of *C. albicans* were not immediately established, primarily because there evolved a variety of genetic fingerprinting methods that were never subjected

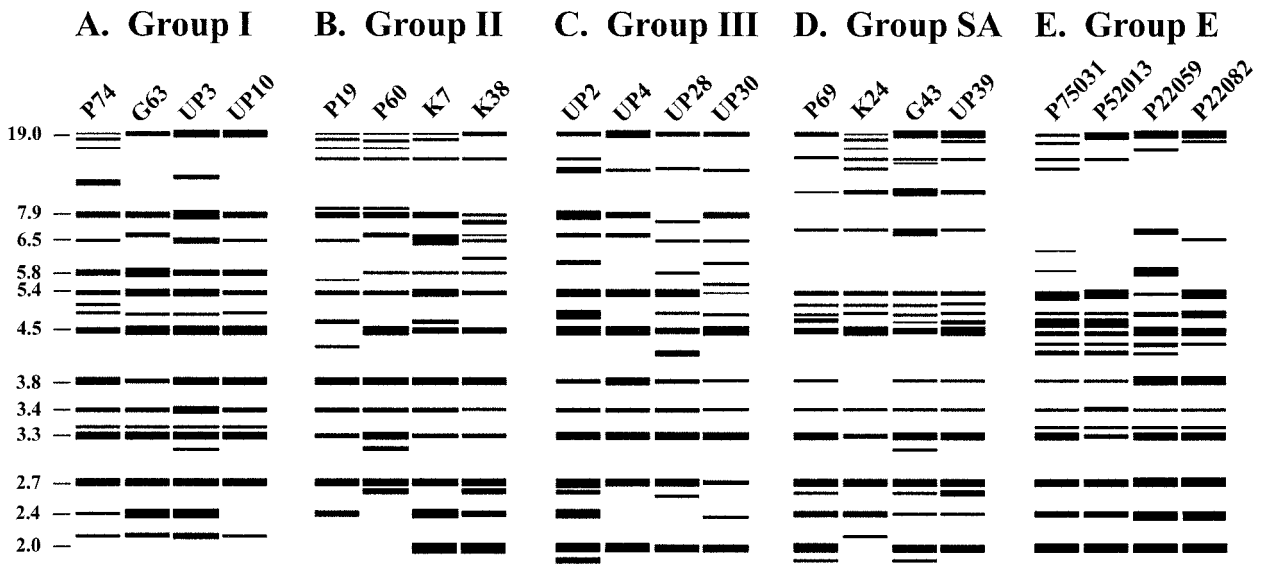

FIG. 6. DENDRON-generated models of Ca3 fingerprinting patterns of randomly selected isolates from group I (A), group II (B), group III (C), group SA (D), and group E (E). Molecular sizes in kilobases of select bands are presented to the left of the models in panel A. Band widths reflect band intensities. Strains are identified above the patterns.

to cross-verification through a comparison of their effectiveness in clustering the same collection of isolates (15). It was not until 1997 that Pujol et al. (9) performed a cluster analysis of 26 unrelated oral isolates from the United States, using four unrelated genetic fingerprinting methods. Three of the methods cross-verified each other and established the three clusters I, II, and III in the test collection (9). This test collection (9) was subsequently used by Blignaut et al. (3) in mixed dendrograms with oral isolates from South Africa to distinguish group I, II, and III isolates in the latter collection. This study revealed a clade specific to South Africa, group SA, that represented half of all isolates colonizing black South Africans and a third of all isolates colonizing white South Africans. No group SA isolates were identified in the 26 U.S. reference isolates, suggesting that the SA clade was specific to South Africa. Since the U.S. reference collection included only 26 isolates, we tested an additional 164 North American bloodstream isolates and extended the study to 46 European and 22 South American isolates. The geographical differences revealed in this study are remarkable.

North America is relatively devoid of group SA isolates. The analysis of an expanded North American collection that included 139 U.S. bloodstream isolates and 25 Canadian bloodstream isolates supported the previous observation by Blignaut et al. (3) that group SA isolates are rare (2%) in North America. The combined data demonstrate that this conclusion is true for bloodstream isolates as well as oral and vaginal isolates. In addition, a comparison of geographical regions of North America revealed that the collection from the U.S. Southwest was devoid of group II isolates, and that the northernmost collection, from Canada, contained the largest proportion of group II isolates (24%). The East Coast and Midwest collections contained intermediate proportions of group II isolates, 13 and 19%, respectively.

Identification of a new European clade. By applying the same method, a new clade was identified in Europe, group E. The group E clade made up 22% of unrelated European isolates collected from nine different countries. In marked contrast, the collections from North America, South America, and South Africa were relatively devoid of group E isolates, making up 1 to 5% of the organisms in the three geographical regions. In the dendrogram of European isolates, the hierarchy of nodes separating groups, beginning with the most deeply rooted, was as follows: (i) group SA, (ii) group III, (iii) group II, and (iv) groups I and E (Fig. 8A). The node hierarchy in the analysis by Blignaut et al. (3) of the South African collections recorded the same hierarchy for groups I, II, III, and SA. In that study, a tree was developed based on nodes and the presence or absence of the IS*1* intron that represented an extension of one proposed by Lott et al. (7). In this tree, first group SA and a progenitor of groups I, II, and III diverged, and then group III and a progenitor of groups I and II emerged. The loss of IS*1* in groups I and II were represented by narrower branches in the tree. We have demonstrated here that roughly half of group E isolates also contain the IS*1* intron, leading to an expanded model of the one proposed by Blignaut et al. (3), in which first, group SA and a progenitor of groups I, II, III, and E diverge, secondly, group III and a progenitor to groups I, II, and E diverge, and finally group E and a progenitor to groups I and II diverge (Fig. 8B). The loss of IS*1* is again represented by a thinner line in the tree (Fig. 8B).

The South American collection is similar to the U.S. Southwest collection. A limited South American collection of 22 isolates was devoid of group II isolates, was nearly devoid of group SA isolates, and contained 5% group E isolates, proportions similar to that of the U.S. Southwest collection. However, while group III was overrepresented in the Southwest collection (51 versus 23% in the South American collection), group I was overrepresented in the South American collection (54 versus 37% in the U.S. Southwest collection). It may be no coincidence that group representation of the U.S. Southwest is similar to that of South America, given geographical proximity.

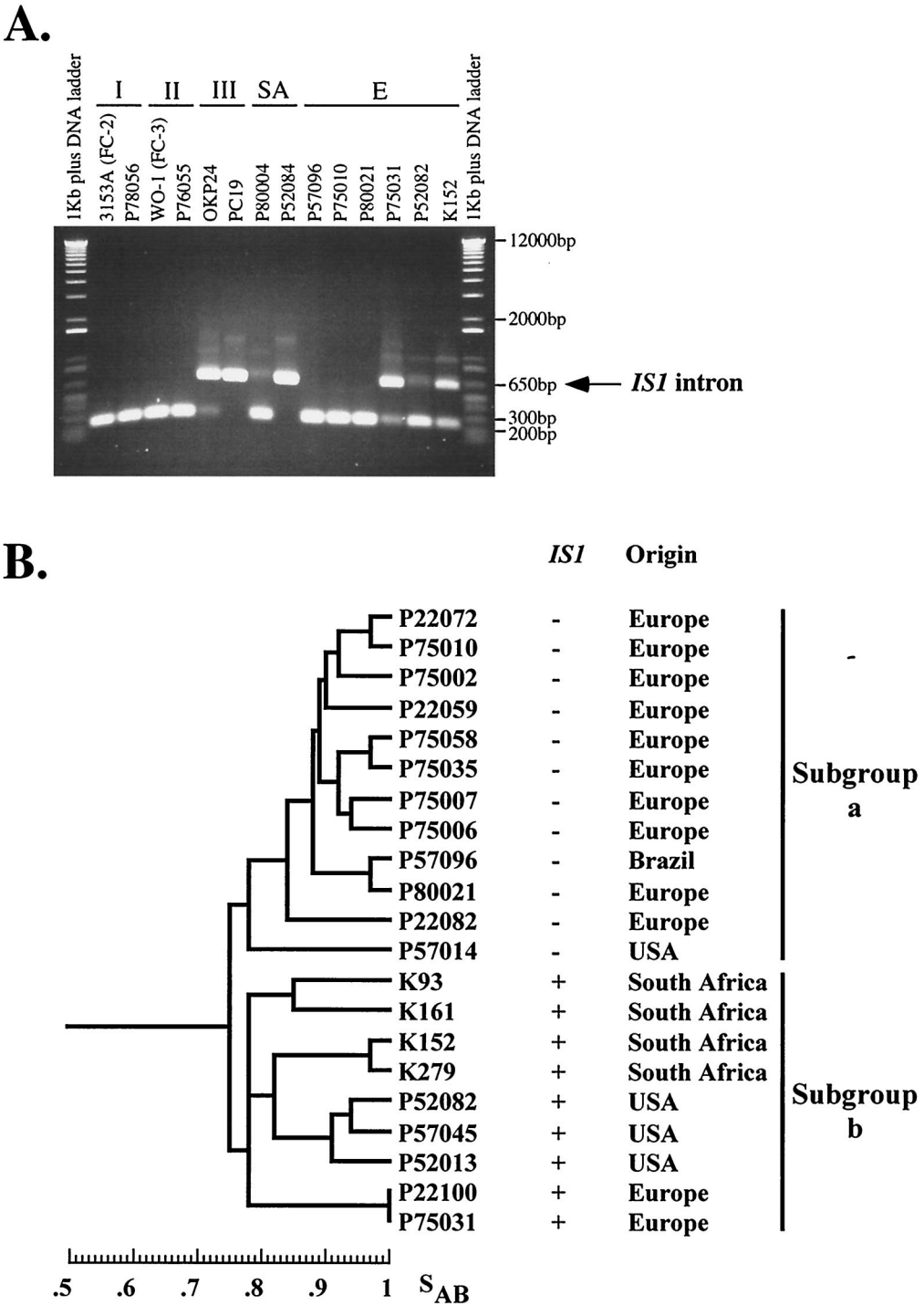

FIG. 7. (A) IS*1* is present in some group E isolates. IS*1* was amplified with IS*1*-specific primers. The IS*1* amplification product was 626 bp. Standards were run in the two outer lanes (1 kb Plus DNA Ladder markers from Invitrogen, Carlsbad, Calif.). Examples of isolates in groups I, II, III, SA, and E are presented. (B) Group E isolates in a dendrogram are scored for the presence of the IS*1* intron. Note that only isolates in subgroup b contain IS*1* introns.

It will be interesting to examine group representation in Central America to see if group II is also missing from that geographical region.

The geographical separation of clades and the possibility of geographically specific reservoirs. The geographical specificity of clades first revealed in the study by Blignaut et al. (3) and now in the present study is surprising. While group SA accounts for half of the isolates colonizing black South Africans and a third of the isolates colonizing white South Africans, it accounts for only 2% of North American isolates, 0% of South American isolates, and only 13% of European isolates. And while group E accounts for 22% of bloodstream isolates in

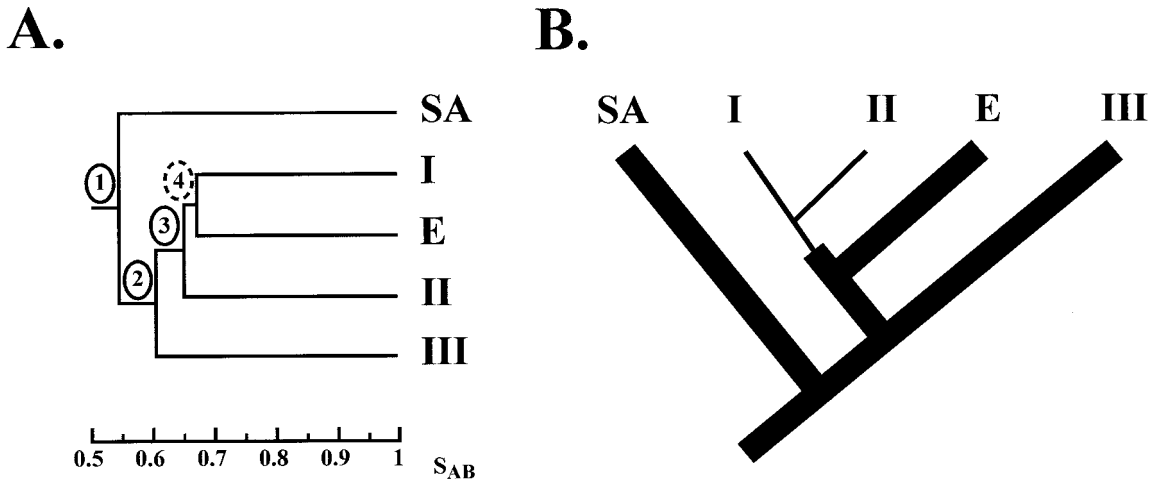

FIG. 8. (A) A model of node hierarchy observed in the dendrogram generated for all European isolates. Because the nodes demarcating groups I, E, and II were very close, the last node, 4, is encircled with a dashed line. (B) A tree representing the evolution of the five major groups of *C. albicans* based both on node hierarchy in dendrograms (panel A) and the presence of IS*1*. This tree is based on the ones developed by Lott et al. (7) and Blignaut et al. (3). The thin branches reflect complete loss of the IS*1* intron from a group.

Europe, it accounts for only 3% of North American isolates, 5% of the South American isolates, and 1% of South African isolates. It is clear that in North America, groups I, II, and III account for the majority of isolates, and even the proportion of outliers is low, demonstrating a low level of diversity. Just as surprising is the absence of group II isolates in the U.S. Southwest collection and in the South American collection. Given the high proportion of individuals of European descent in North America and in the white population of South Africa, one cannot readily attribute the differences to race. In addition, given the fact that *C. albicans* is a common commensal and given the mobility of host populations through air travel, one is hard pressed to explain why the most successful strains don't take over all geographical locales. One might argue that there is no most successful strain and that the differences are a consequence of random genetic drift resulting from partially isolated populations. However, it is more likely that the answer lies in the differences in geographies. If *C. albicans* colonization is continuously refurbished from a nonhuman reservoir, either animal or environmental, that selects for strains, and hence clades, one can explain the extraordinary differences in the proportions of clade representation in the different geographical locales. The challenge that now faces us is to consider all explanations for the clear geographical differences revealed in these studies and then test these explanations. If differences in reservoirs emerge as a viable explanation, what or where are these reservoirs and what distinguishes them among North America, Europe, South America, and South Africa? Are there sharp geographical boundaries in group representation, or smooth transitions? How many more distinct clades of *C. albicans* exist in as yet untested geographical locales across the world? And finally, what phenotypic differences distinguish these clades? We strongly recommend that in pursuit of the answers to these questions, the Ca3-based fingerprinting methods used in the study by Blignaut et al. (3) and in the present study continue to be employed so that the data from each successive study can be pooled.

ACKNOWLEDGMENTS

We thank S. Jolie for participation in the initial stages of the analysis.

This research was supported by a grant from Pfizer Corporation to M.P. and D.R.S. and NIH grant AI2392 to D.R.S.

REFERENCES

1. **Anderson, J., T. Srikantha, B. Morrow, S. H. Miyasaki, T. C. White, N. Agabian, J. Schmid, and D. R. Soll.** 1993. Characterization and partial nucleotide sequence of the DNA fingerprinting probe Ca3 of *Candida albicans.* J. Clin. Microbiol. **31:**1472–1480.
2. **Anderson, J. B., C. Wickens, M. Khan, N. Federspiel, T. Jones, and L. M. Kohn.** 2001. Infrequent genetic exchange and recombination in the mitochondrial genome of *Candida albicans*. J. Bacteriol. **183:**865–872.
3. **Blignaut, E., C. Pujol, S. Lockhart, S. Joly, and D. R. Soll.** 2002. Ca3 fingerprinting of *Candida albicans* isolates from human immunodeficiency virus-positive and healthy individuals reveals new clade in South Africa. J. Clin. Microbiol. **40:**826–836.
4. **Hasenclever, H. F., and W. O. Mitchell.** 1961. Antigenic studies of *Candida.* I. Observations of two antigenic groups in *Candida albicans*. J. Bacteriol. **82:**570–573.
5. **Hasenclever, H. F., and W. O. Mitchell.** 1961. Antigenic studies of *Candida.* III. Comparative pathogenicity of *Candida albicans* group A, group B, and *Candida stellatoidea*. J. Bacteriol. **82:**578–581.
6. **Lockhart, S., J. J. Fritch, A. S. Strudevant Meier, K. Schroppel, T. Srikantha, R. Galask, and D. R. Soll.** 1995. Colonizing populations of *Candida albicans* are clonal in origin but undergo microevolution through C1 fragment reorganization as demonstrated by DNA fingerprinting and C1 sequencing. J. Clin. Microbiol. **33:**1501–1509.
7. **Lott, T. J., B. P. Holloway, D. A. Logan, R. Fundyga, and J. Arnold.** 1999. Towards understanding the evolution of the human commensal yeast *Candida albicans*. Microbiology **145:**1137–1143.
8. **Pfaller, M. A., D. J. Diekema, R. N. Jones, H. S. Sader, A. C. Fluit, R. J. Hollis, S. A. Messer, and The Sentry Participant Group.** 2001. International surveillance of bloodstream infections due to *Candida* species: frequency of occurrence and in vitro susceptibilities to fluconazole, ravuconazole, and voriconazole of isolates collected from 1997 through 1999 in the SENTRY Antimicrobial Surveillance Program. J. Clin. Microbiol. **39:**3254–3259.
9. **Pujol, C., S. Joly, S. R. Lockhart, S. Noel, M. Tibayrenc, and D. R. Soll.** 1997. Parity among the randomly amplified polymorphic DNA method, multilocus enzyme electrophoresis, and Southern blot hybridization with the moderately repetitive DNA probe Ca3 for fingerprinting *Candida albicans*. J. Clin. Microbiol. **35:**2348–2358.
10. **Pujol, C., S. Joly, B. Nolan, T. Srikantha, and D. R. Soll.** 1999. Microevolutionary changes in *Candida albicans* identified by the complex Ca3 fingerprinting probe involve insertions and deletions of the full-length repetitive sequence RPS and specific genomic sites. Microbiology **145:**2635–2646.
11. **Pujol, C., J. Reynes, F. Renaud, M. Raymond, M. Tibayrenc, F. J. Ayala, F. Janbon, M. Mallie, and J. M. Bastide.** 1993. The yeast *Candida albicans* has

a clonal mode of reproduction in a population of infected human immunodeficiency virus-positive patients. Proc. Natl. Acad. Sci. USA **90:**9456–9459.
12. **Sadhu, C., M. J. McEachern, E. P. Rustchenko-Bulgac, J. Schmid, D. R. Soll, and J. B. Hicks.** 1991. Telomeric and dispersed repeat sequences in *Candida* yeasts and their use in strain identification. J. Bacteriol. **173:**842–850.
13. **Scherer, S., and D. A. Stevens.** 1987. Application of DNA fingerprinting methods to epidemiology and taxonomy of *Candida* species. J. Clin. Microbiol. **25:**675–679.
14. **Schmid, J., E. Voss, and D. R. Soll.** 1990. Computer-assisted methods for assessing strain relatedness in *Candida albicans* by fingerprinting with the moderately repetitive sequence Ca3. J. Clin. Microbiol. **28:**1236–1243.
15. **Soll, D. R.** 2000. The ins and outs of DNA fingerprinting the infectious fungi. Clin. Microbiol. Rev. **13:**332–370.
16. **Warren, N. G., and K. C. Hazen.** 1995. *Candida*, *Cryptococcus*, and other yeasts of medical importance, p. 723–737. *In* P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover, and R. H. Yolken (ed.), Manual of clinical microbiology, 6th ed. American Society for Microbiology, Washington, D.C.