Title
Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis.

Journal
Proceedings of the National Academy of Sciences of the United States of America, 90(4)

ISSN
0027-8424

Authors
Tibayrenc, M
Neubauer, K
Barnabé, C
et al.

Publication Date
1993-02-15

DOI
10.1073/pnas.90.4.1335

Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 1335–1339, February 1993
Evolution

Genetic characterization of six parasitic protozoa: Parity between random-primer DNA typing and multilocus enzyme electrophoresis

(phylogeny/clonal theory/*Trypanosoma*/*Leishmania*/*Plasmodium*)

MICHEL TIBAYRENC*†, KATJA NEUBAUER, CHRISTIAN BARNABÉ*, FRANÇOISE GUERRINI*, DOUGLAS SKARECKY, AND FRANCISCO J. AYALA

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92917

Contributed by Francisco J. Ayala, November 12, 1992

ABSTRACT **We have assayed genetic polymorphisms in several species of parasitic protozoa by means of random amplified polymorphic DNA (RAPD). One goal was to ascertain the suitability of RAPD markers for investigating genetic and evolutionary problems, particularly in organisms, such as the parasitic protozoa, unsuitable for traditional methods of genetic analysis. Another goal was to test certain hypotheses concerning *Trypanosoma cruzi*, and other protozoa, that have been established by multilocus enzyme electrophoresis. The RAPD results corroborate the hypothesis that the population structure of *T. cruzi* is clonal and yield a phylogeny of the clonal lineages in agreement with the one obtained by enzyme electrophoresis. This parity between the two sets of results confirms that RAPD markers are reliable genetic markers. The RAPD markers are also suitable for reconstructing species phylogenies and as diagnostic characters of species and subspecific lineages. The number of DNA polymorphisms that can be detected by the RAPD method seems virtually unlimited, since the number of primers can be increased effectively at will. The RAPD method is well suited for investigating genetic and evolutionary questions in certain organisms, because it is cost effective and demands no previous genetic knowledge about the organism.**

The recently developed technique of random amplified polymorphic DNA (RAPD) provides an effective method for obtaining genetic markers in all sorts of organisms (1–4). The RAPD method identifies polymorphisms that are detected as DNA fragments, amplified by the *Taq* polymerase chain reaction (PCR), that are present in one but not another individual or strain. Short oligonucleotides (about 10-mer) of arbitrary nucleotide sequence are often used as amplifying primers, although nonrandom longer primers have proved useful for certain purposes (4).

The RAPD method holds considerable promise for population genetic and evolutionary studies (5) because (*i*) no previous sequence information is needed in the target organism and hence all sorts of organisms are accessible, (*ii*) many markers can readily be identified, as is required for the reconstruction of phylogenetic history, and (*iii*) the effort and cost involved are modest so that many individuals can be assayed, which is usually necessary for investigating population genetic problems.

The validity of the method in broad practice remains, however, to be established. Questions arise as to the consistency of the results, the genetic significance of the amplified DNA sequences, and even the possibility of artifactual outcomes (6). In situations where genetic crosses are not possible, either because the organisms belong to different species or because sexual reproduction is absent (or not feasible experimentally), these issues cannot be resolved by traditional Mendelian methods; they can be investigated only by indirect methods.

We summarize a RAPD study of several parasitic protozoa. The results are consistent with data previously obtained by multilocus enzyme electrophoresis (MLEE). The genetic and evolutionary significance of the RAPD method is thus established in this case. Moreover, the data corroborate (*i*) that *Trypanosoma cruzi* has a clonal mode of propagation, as we have previously established (7, 8), and (*ii*) phylogenetic relationships among the parasites that are consistent with previous knowledge. These two results are of epidemiological and medical significance.

MATERIALS AND METHODS

We used the twenty 10-mer primers (kit A; obtained from Operon Technologies, Alameda, CA) listed in Table 1. Conditions were as published (1) and recommended by Boehringer Mannheim, supplier of the *Taq* DNA polymerase. The PCR-amplified DNA fragments were separated in 1% agarose gels and stained with ethidium bromide. MLEE was carried out for 12–22 gene loci, following published methods (7–9).

T. cruzi, the agent of Chagas disease, was represented by 24 stocks, listed in Fig. 1. They derived from four countries (Brazil, 10 stocks; Bolivia, 6; Chile, 5; Venezuela, 3) and included 14 stocks of clonets 19, 20, and 39 (a "clonet" designates, in a clonal species, all isolates that are identical for a particular set of genetic markers; see refs. 11 and 12) that have wide geographic distribution (8), plus 10 stocks representing geographically restricted clonets. Two additional *Trypanosoma* taxa from South America and two from Africa (*T. brucei brucei* and *T. brucei gambiense*) were each represented by one stock, and so were five species of South American *Leishmania*. Finally, two commonly available laboratory stocks of *Plasmodium falciparum*, the agent of malignant malaria, were included, one from Central America and one from Africa (Table 1).

Phylogenetic analyses were based on Jaccard's distance (13), which simply measures the proportion of band mismatches between pairs of stocks. Each RAPD gel band is coded with a number, starting with 1 for the slowest band (largest DNA fragment). Each stock is thus represented by a set of numbers for each primer. Jaccard's distance is $D_{ij} = 1 - [C/(2N - C)]$, where C is the number of bands in common

Abbreviations: RAPD, random amplified polymorphic DNA; MLEE, multilocus enzyme electrophoresis; UPGMA, unweighted pair-group method with arithmetic averages.

*Permanent address: Unité Mixte de Recherche, Centre National de la Recherche Scientifique–Institut Français de Recherche pour le Développement en Coopération No. 9926, Génétique Moléculaire des Parasites et des Insectes Vecteurs, BP 5045, 34032 Montpellier, Cédex 01, France.

†To whom reprint requests should be addressed at permanent address.

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

Table 1. Random primer DNA amplification in six parasitic protozoa

	T. cruzi	*T. cruzi marenkellei*	*T. rangeli*	*T. brucei*	*Leishmania*	*P. falciparum*
Stocks (*n*)	24	1	1	2	5	2
Enzyme loci (*n*)	22	22	22	18	15	12
Primers						
1. CAGGCCCTTC	+	+	+	+	+,C	+
2. TGCCGAGCTG	+	+	+	+,M	+	−
3. AGTCAGCCAC	+	+	+	+	+	−
4. AATCGGGCTG	+	+	+	+,C	+,C	+,C
5. AGGGGTCTTG	+,C	+	+	+,C	+,C	+
6. GGTCCCTGAC	−	+	+	−	+	+
7. GAAACGGGTG	+,C	+	+	+,C	+,C	+
8. GTGACGTAGG	+,M	+	+	+,C	+	+,C
9. GGGTAACGCC	+	+	+	+,C	+	+,C
10. GTGATCGCAG	+,C,S	+	+	+,C	+,C	+,C
11. CAATCGCCGT	+	−	−	−	+	−
12. TCGGCGATAG	+	+	+	+	+	−
13. CAGCACCCAC	−	+	+	−	−	−
14. TCTGTGCTGG	−	−	−	−	−	−
15. TTCCGAACCC	+,C,S	+	+	−	+,C	+
16. AGCCAGCGAA	−	+	+	+	+	−
17. GACCGCTTGT	+,C,S	+	+	+,M	−	+
18. AGGTGACCGT	+	−	−	+,M	+	+
19. CAAACGTCGG	+,C,S	−	−	−	−	+
20. GTTGCGATCC	+,M	+	+	+,M	+	+

The 20 oligonucleotide primers were from Operon Technologies, kit A. −, Unsatisfactory; +, satisfactory amplified band pattern; C, pattern used for calculating Jaccard's distances; M, monomorphic pattern; S, synapomorphic pattern. Stocks of *T. cruzi* are listed in Fig. 1 legend. *T. brucei* stocks: one each of *T. brucei brucei* and *T. brucei gambiense*. *Leishmania*: one each of *L. amazonensis, L. braziliensis, L. guyanensis, L. infantum,* and *L. mexicana*; *Plasmodium falciparum*: two commonly available laboratory stocks, one from Honduras and one from Gambia (SGE 1).

between the *i* and *j* stocks and *N* is the total number of different bands in both stocks. The same method was used for generating the MLEE distance matrices. Dendrograms were derived from the distance matrices by using the UPGMA (unweighted pair-group method with arithmetic averages) algorithm (14).

RESULTS AND DISCUSSION

Most but not all of the 20 tested primers yielded gel patterns that were consistent from gel to gel and from sample to sample (Table 1). Some primers yielded patterns that were fuzzy or inconsistent for some (e.g., A11, A12) or all (A14) taxa. Two primers yielded gel patterns that were simple and identical ("monomorphic") for all 24 *T. cruzi* stocks but variable among taxa (Table 1 and Fig. 2). Most primers yielded patterns that varied among the *T. cruzi* stocks (Fig. 3), as well as among taxa. The gel patterns used for calculating the Jaccard distances are labeled C in Table 1; other patterns either are monomorphic in a given species or are excessively complex and have been ignored for the present purposes.

Reconstruction of Phylogeny. Fig. 1 shows two UPGMA dendrograms for the 24 *T. cruzi* stocks, one based on the RAPD data and the other on the MLEE data. The topologies of the two dendrograms are similar. This agreement supports the inference that both methods reveal the same biological reality—namely, the extent of genetic differentiation among the organisms (and hence their phylogenetic relationships). The RAPD technique is thereby validated for these organisms as a method for detecting DNA polymorphisms that mark distinctive genetic makeups. It is worth noticing that the data set used in Fig. 1 for the RAPD dendrogram derives from only six primers (see Table 1; two primers gave *T. cruzi* gel patterns that were monomorphic and hence did not yield any phylogenetic information for this parasite; the complex patterns manifested by the remaining 12 primers were eliminated for calculating distances without any consideration of the MLEE data).

We have examined the correspondence between the RAPD and MLEE results by calculating the correlation coefficient between the two sets of distances for the 276 pairwise comparisons among the 24 *T. cruzi* stocks. The correlation coefficient is $r = 0.924$, which is statistically highly significant [$P < 0.001$, conservatively assuming only 22 degrees of freedom (df)]. The correlation between the two sets of distances is also statistically significant within each subset defined by the two major clades shown in Fig. 1 ($r = 0.734$ for the 78 pairwise comparisons between the 13 stocks in the top cluster, $P < 0.05$ for 11 df; $r = 0.886$ for the 55 comparisons between the 11 stocks in the lower clade, $P < 0.01$ for 9 df).

The parity between the RAPD and MLEE results can also be tested with the set of five *Leishmania* species. The Jaccard distances are shown in Table 2; the corresponding dendrograms are in Fig. 4. The correspondence between the two sets of data is apparent, although the data set is small and the association of the *infantum* lineage with the *amazonensis/mexicana* clade rather than with *braziliensis/guyanensis* is not statistically significant for either dendrogram.

The RAPD patterns for the two *T. brucei* stocks (subspecies *brucei* and *gambiense*) did not differ for four primers. The data for six additional primers with readily interpretable patterns yielded a Jaccard distance of 0.63; the MLEE distance based on 18 loci is 0.39 (F. Mathieu-Daudé and M.T., unpublished data). The two *P. falciparum* stocks are genetically quite different from each other, with distances of 0.98 and 0.36 for the RAPD and MLEE data, respectively (S. Ben Abderrazak and M.T., unpublished data).

Clonal Population Structure. We have earlier shown that *T. cruzi* has a clonal population structure (7, 8, 15, 16). A telltale sign of clonal propagation is the association between unrelated polymorphic markers, an especially striking instance of linkage disequilibrium (8, 15–18). We have no knowledge of which DNA polymorphisms are amplified by the RAPD primers (although this could be investigated by cloning and sequencing DNA isolated from gel bands). But one would not expect any substantial overlap between the DNA sequences expressed in the MLEE and RAPD polymorphisms, or

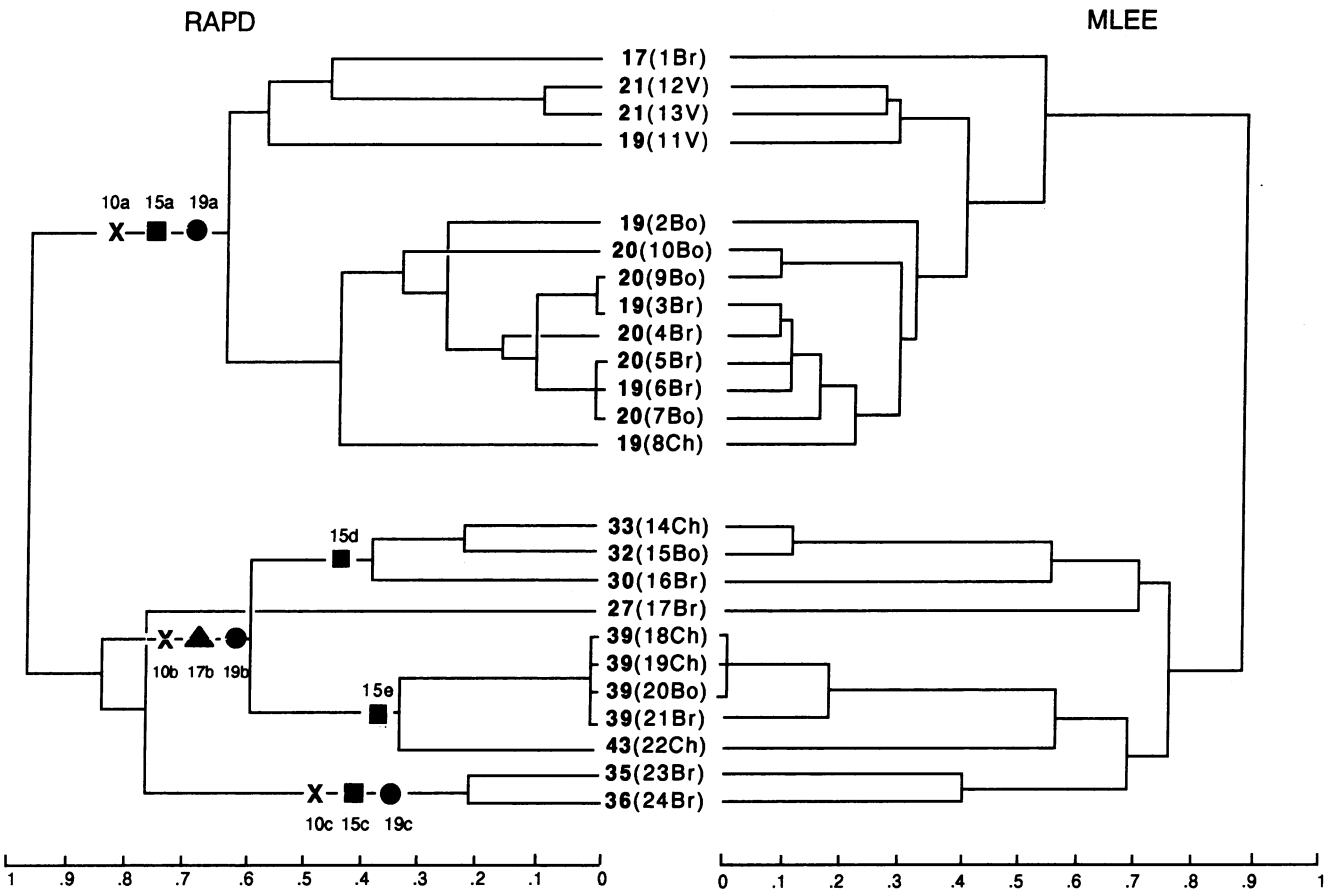

FIG. 1. Dendrograms of the 24 *T. cruzi* stocks derived from genetic distances obtained by MLEE (*Right*) and RAPD (*Left*) methods. The first number identifying each stock is the clonet number to which the stock belongs (ref. 8; the term "zymodeme" rather than "clonet" is used in ref. 8); the numbers in parentheses are the identification numbers used in the present study; the letters refer to the countries of origin: Bolivia (Bo), Brazil (Br), Chile (Ch), and Venezuela (V). (Notice that some stocks, such as nos. 18–21, that appeared identical at all enzyme loci in ref. 8—and hence were classified within the same clonet, no. 39 in the example—are heterogeneous when more loci are assayed; thus stock 21 has now a MLEE pattern somewhat different from those of nos. 18–20.) Information about 17 stocks can be found in ref. 8 (we give the stock number used in the present study on the left of the equal sign and the number used in ref. 8 on the right): 1 = 23, 2 = 50, 3 = 28, 4 = 64, 5 = 65, 6 = 29, 11 = 34, 12 = 68, 13 = 67, 15 = 81, 16 = 78, 17 = 75, 20 = 94, 21 = 110, 22 = 121, 23 = 90, and 24 = 91. For the other stocks: 7 = SO34 cl4, from Bolivia, isolated from *Triatoma infestans*; 8 = LGN, Chile, human; 9 = P11 cl3, Bolivia, human; 10 = P209 cl1, Bolivia, human; 14 = CBB, Chile, human; 18 = MN cl2, Chile, human; 19 = NR cl9, Chile, human. Notice that the top clade, comprising 13 stocks, approximately corresponds to zymodeme I in ref. 10. The symbols on the RAPD dendrogram represent synapomorphic characters; these are labeled with a number that indicates the primer used and a letter that refers to a particular band pattern.

between the DNA sequences amplified by different RAPD primers. Therefore, the correspondence between the RAPD and MLEE data evinces linkage disequilibrium between these two data sets (as well as between the different RAPD polymorphisms) and thus confirms that *T. cruzi* has a clonal population structure, as we have previously established on the basis of MLEE data (7, 8) and corroborated with restriction fragment length polymorphism (RFLP) analysis of kinetoplastid DNA (17).

If we infer that different genetic markers are amplified by different RAPD primers, we can apply the tests we have earlier proposed (15, 16) to examine nonrandom association between loci. The results of these tests are statistically highly significant (as might be expected in view of the parity between the RAPD and MLEE distance matrices and dendrograms). The tests have been described (15, 16). The probability values obtained are *d1* (combinatorial probability of observing the most common multilocus genotype as many or more times as found in the sample), $P = 8 \times 10^{-11}$; *d2* (Monte Carlo probability of sampling as many or more individuals of any genotype as observed of the most common genotype in the sample), $P < 0.0001$; *e* (probability of sampling as few or fewer different genotypes as observed), $P < 0.0001$; and *f* (probability of observing as large or larger genetic disequilibrium as found), $P < 0.0001$. We have also calculated the statistic $\mathrm{Var}(d_{ij})$, which measures multilocus associations by calculating the number of allelic mismatches between $n(n-1)/2$ pairs of individuals sampled (19); statistical significance is estimated as the proportion of samples with a variance greater than the observed value, $P < 0.0001$. (This test, as well as tests of *d2*, *e*, and *f*, is based on computer simulations, each with 10,000 runs.)

A striking state of affairs discovered earlier (7, 8) is the occurrence of multilocus genotypes that are repeatedly found in distant localities and at different times. We have referred to these overrepresented genotypes as "ubiquitous" and called attention to them because they evince clonal propagation that persists on the evolutionary scale and also because potentially they are of great epidemiological significance (11, 12, 15, 16). We have included in the RAPD analysis three stocks (nos. 18–20 in Table 1) representative of clonet 39 (ref. 8) that exhibit identical genotypes at the 22 enzyme loci assayed. One of these stocks was collected in Bolivia in 1981 from a triatomine insect vector; the other two were collected from humans in Chile 600 km apart, one in 1987, the other in 1988. The patterns obtained with all 18 RAPD primers (two primers give fuzzy results with *T. cruzi* stocks) are identical for these three stocks.

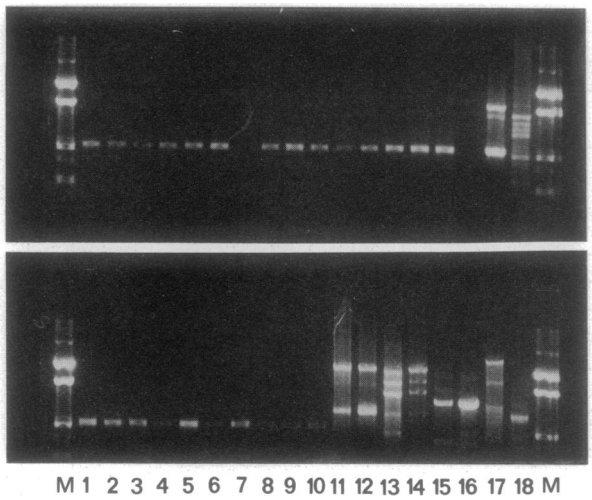

FIG. 2. RAPD patterns obtained with primer 8. (*Upper*) From left to right, lanes 1–15 are *T. cruzi* stocks 1, 2, 3, 6, 8, 11, 4, 5, 9, 10, 7, 12, 13, 24, and 23; lane 16, blank; lane 17, *T. cruzi marenkellei*; lane 18, *T. rangeli*. (*Lower*) Lanes 1–10 are *T. cruzi*, lane 1 is stock Bug2145 from Brazil, clonet 19; lanes 2–5 are stocks 21, 18, 19, and 20; lane 6 is 503 cl5 from Bolivia, clonet 39; lanes 7–10 are stocks 15, 14, 16, and 22; lane 11, *T. brucei brucei*; lane 12, *T. brucei gambiense*; lane 13, *Leishmania braziliensis*; lane 14, *L. guyanensis*; lane 15, *L. mexicana*; lane 16, *L. amazonensis*; lane 17, *L. infantum*; lane 18, *P. falciparum*; lanes M, size markers.

Species-Diagnostic Patterns. All RAPD primers yield patterns that differ between species and thus could serve as species-diagnostic traits (if one assumes, of course, that the between-species differences would persist with larger samples). But the extensive polymorphisms observed within each species for most RAPD primers preclude their practical use for species diagnosis. There are, however, primers that yield monomorphic patterns within a species and thus can readily be used for species diagnosis, which may be useful for epidemiological and other purposes. For example, two primers (nos. 8 and 20 in Table 1) are monomorphic in *T. cruzi* and thus could be used in epidemiological practice for differentiating *T. cruzi* from *T. rangeli*, a species that infects humans (as well as triatomine vectors) but is thought to be nonpathogenic. Discrimination between isolates of the two species is not always readily successful with currently used methods. (We have observed simple monomorphic patterns also in *T. brucei*, but their potential diagnostic application will depend on confirming their monomorphic status by examining more stocks.)

Synapomorphic Patterns. Several RAPD primers yielded synapomorphic patterns—i.e., homologous patterns that are shared by all lineages in a clade and are inferred to have been present in their nearest common ancestor but not in earlier ancestors or in lineages outside this clade. The synapomorphic patterns are indicated in Fig. 1 by a number that refers to a particular primer and a letter that refers to a particular band pattern. Thus, primer 10 yields a pattern (10a) that is

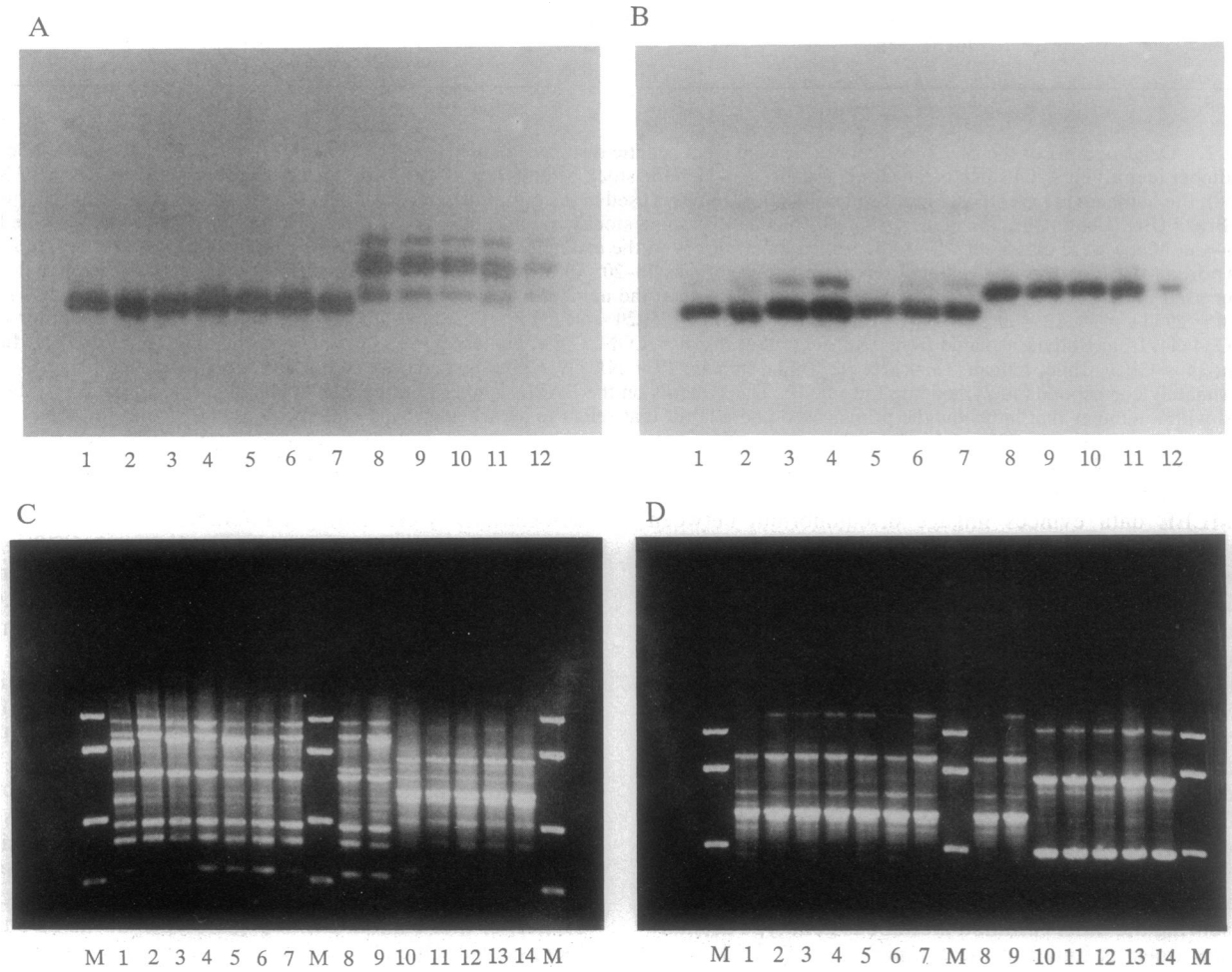

FIG. 3. MLEE (*Upper*) and RAPD (*Lower*) gel patterns for *T. cruzi* stocks. (*A*) Glucose-6-phosphate isomerase (EC 5.3.1.9). Lanes 1–4, clonet 19 (stocks 2, 3, and 6, and an isolate from *Triatoma spinolai*, from Chile); lanes 5–7, clonet 20 (stocks 4, 5, and 9); lanes 8–12, clonet 39 (stocks 21, 18, 19, and 20). (*B*) Glutamic oxaloacetic transaminase (EC 2.6.1.1). Lane loading as in *A*. (*C*) Primer 7. Lanes 1–7, as in *A*; lanes 8 and 9, clonet 20 (stocks 10 and 7); lanes 10–14, as in *A* lanes 8–12; lanes M, size markers. (*D*) Primer 10. Lane loading as in *C*.

Table 2. Genetic distances between five species of *Leishmania* based on RAPD (below diagonal) and MLEE (above diagonal) data

	braziliensis	*guyanensis*	*amazonensis*	*mexicana*	*infantum*
braziliensis		0.47	0.98	0.98	0.85
guyanensis	0.38		0.97	0.98	0.91
amazonensis	0.96	0.91		0.57	0.89
mexicana	0.92	0.93	0.81		0.91
infantum	0.96	0.93	0.90	0.84	

Distances were estimated according to ref. 13.

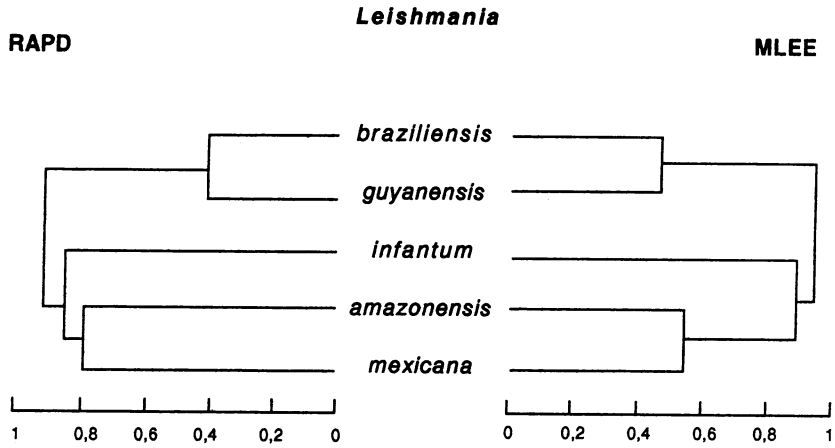

FIG. 4. Dendrograms of five *Leishmania* species, derived from genetic distances obtained by MLEE (*Right*) and RAPD (*Left*).

synapomorphic for the top clade in Fig. 1, encompassing 13 stocks, another pattern (10b) that is synapomorphic for a clade comprising 8 stocks, and one more pattern (10c) synapomorphic for the bottom 2 stocks. Primers 15, 17, and 19 also yield synapomorphic patterns. Synapomorphies serve not only as phylogenetic markers but could also be used for clinical or epidemiological purposes. The evolutionary propinquity of the lineages within a clade may often implicate distinctive biological attributes of medical consequence. In practice, the use of RAPD synapomorphies for identifying stocks can be accomplished not only by RAPD assay, but also by means of DNA probes specific for the synapomorphic DNA fragments.

Concluding Remarks. We have shown that the RAPD method detects genetic markers useful for reconstructing evolutionary history, population genetic analysis, species diagnosis, and clade membership. The number of different primers that can be used is virtually unlimited. Most of the 20 primers we have tested yield informative markers for any one protozoan species. If a similar rate of success obtains with other primers and other organisms, the RAPD method opens up the possibility of securing as many genetic markers as may be wanted—and with a moderate investment of time and resources. Traditional genetic analysis of the detected polymorphisms is possible for many organisms. Moreover, the DNA fragments can be isolated, sequenced, and characterized.

We are grateful to J. Welsh and M. McClelland (California Institute of Biological Research, La Jolla, CA) for technical advice; B. Bastrenta and F. Brenière for technical support; A. Solari for providing three Chilean stocks; F. Mathieu-Daudé and S. Ben Abderrazak for sharing unpublished data with us; and D. L. Hartl and R. Milkman for valuable comments about the manuscript. This work was supported by World Health Organization Training Diseases Research (TDR) Chagas Grant 880190, World Health Organization TDR Epidemiology of African Trypanosomiasis (EP/AF) 920172, North Atlantic Treaty Organization Grant 911040, a grant from the Fondation pour la Recherche Medicale, and National Institutes of Health Grant GM42397.

1. Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A. & Tingey, S. V. (1990) *Nucleic Acids Res.* **18,** 6531–6535.
2. Welsh, J. & McClelland, M. (1990) *Nucleic Acids Res.* **18,** 7213–7218.
3. Welsh, J. & McClelland, M. (1991) *Nucleic Acids Res.* **19,** 861–866.
4. Welsh, J., Petersen, C. & McClelland, M. (1991) *Nucleic Acids Res.* **19,** 303–306.
5. Hedrick, P. (1992) *Nature (London)* **355,** 679–680.
6. Riedy, M. F., Hamilton, W. J. & Aquadro, C. F. (1992) *Nucleic Acids Res.* **20,** 918.
7. Tibayrenc, M., Ward, P., Moya, A. & Ayala, F. J. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 115–119.
8. Tibayrenc, M. & Ayala, F. J. (1988) *Evolution* **42,** 277–292.
9. Ben Abderrazak, S., Guerrini, F., Mathieu-Daudé, F., Truc, P., Neubauer, K., Lewicka, K., Barnabé, C. & Tibayrenc, M. (1993) *Protocols in Molecular Parasitology* (Humana, Clifton, NJ), in press.
10. Miles, M. A., Souza, A., Pova, M., Shaw, J. J., Lainson, R. & Toyé, P. J. (1978) *Nature (London)* **272,** 819–821.
11. Tibayrenc, M. & Ayala, F. J. (1991) *Parasitol. Today* **7,** 228–232.
12. Tibayrenc, M., Kjellberg, F. & Ayala, F. J. (1991) *BioScience* **41,** 767–774.
13. Jaccard, P. (1908) *Bull. Soc. Vaudoise Sci. Nat.* **44,** 223–270.
14. Sneath, P. H. A. & Sokal, R. R. (1973) *Numerical Taxonomy* (Freeman, San Francisco).
15. Tibayrenc, M., Kjellberg, F. & Ayala, F. J. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2414–2418.
16. Tibayrenc, M., Kjellberg, F., Arnaud, J., Oury, B., Brenière, S. F., Dardé, M. L. & Ayala, F. J. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 5129–5133.
17. Tibayrenc, M. & Ayala, F. J. (1987) *C. R. Seances Acad. Sci., Ser. 3* **304,** 89–92.
18. Miller, R. D. & Hartl, D. L. (1986) *Evolution* **40,** 1–12.
19. Sved, J. (1968) *Genetics* **59,** 543–563.