Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites

RIVA P. OLIVEIRA*, NATALIA E. BROUDE†, ANDREA M. MACEDO*, CHARLES R. CANTOR†, CASSANDRA L. SMITH†, AND SÉRGIO D. J. PENA*‡

*Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, 31270-901 Brazil; †Center for Advanced Biotechnology and Department of Biomedical Engineering, Boston University, Boston, MA 02215

Contributed by Charles R. Cantor, December 23, 1997

ABSTRACT We describe here the identification of eight polymorphic microsatellite loci with (CA)_n repeats in the *Trypanosoma cruzi* genome based on the affinity capture of fragments using biotinylated (CA)_12 attached to streptavidin-coated magnetic beads. The presence of two peaks in PCR amplification products from individual clones confirmed that *T. cruzi* is diploid. Hardy–Weinberg and linkage disequilibrium analyses suggested that sexual reproduction is rare or absent and that the population structure is clonal. Several strains, especially those isolated from nonhuman sources, showed more than two alleles in many loci demonstrating that they were multiclonal. The phylogenetic analysis of *T. cruzi* based on microsatellites revealed a great genetic distance among strains, although the strain dispersion profile in the Wagner network was in general agreement with the species dimorphism found by PCR amplification of the divergent region of the rRNA 24Sα gene.

The protozoan *Trypanosoma cruzi* is the causative agent of Chagas’ disease, which is estimated to affect 18 million people in Latin America (1). The clinical features of the chronic phase of the disease are quite variable, taking an asymptomatic indeterminate course or evolving as chronic cardiomyopathy and/or gastrointestinal pathology. It is believed that such clinical variability is the result of a complex interplay of environmental and genetic factors from both host and parasite (reviewed in ref. 2). Thus, studies of the genetic structure of natural populations of *T. cruzi* will be important for a better understanding of the pathogenesis of Chagas’ disease.

Characterization of natural populations of *T. cruzi* stocks, initially performed using a small number of isoenzyme markers, suggested a classification into three distinct groups that were called zymodemes (3, 4). Afterward, more extensive studies with 15 enzyme markers in *T. cruzi* clones from the whole ecogeographical range of the parasite and from various hosts increased significantly the number of natural isoenzyme strains or “clones” (5). Although the isoenzyme studies provided convincing evidence that *T. cruzi* is diploid, there were significant deviations from Hardy–Weinberg expectations and strong linkage disequilibrium, suggesting that sexual reproduction is rare or absent and that the population structure is that of independently evolving clonal lineages (reviewed in ref. 6 and 7).

The advent of DNA studies based on multilocus markers such as DNA fingerprinting (8), random-amplified polymorphic DNA (9, 10), simple sequence repeat-PCR (11), and molecular karyotyping (12) showed a much greater level of intraspecific genetic variability of *T. cruzi* than isoenzymes, suggesting an absolute genetic individuality of clonal lineages.

More recently, single locus analysis of the ribosomal RNA genes (LSU-rDNA) and mini-exon genes (13, 14) has suggested a division of *T. cruzi* strains into two major clusters, which probably represent subspecies (14).

The study of microsatellite markers has had an evolutionary impact in human molecular genetics (see, for instance, ref. 15). We reasoned that examination of microsatellite markers from the *T. cruzi* genome should likewise provide genetic information about the parasite. We report here the isolation of eight (CA)_n polymorphic microsatellites from *T. cruzi* by using a magnetic capture affinity technique. Analysis of strains and clones of *T. cruzi* with these microsatellites provided valuable new information about the population structure of the parasite.

MATERIALS AND METHODS

Parasites and Nucleic Acids Isolation. The strains and clones of *T. cruzi* examined are listed in Table 1. Epimastigote forms were grown in liver infusion tryptose medium. The parasites were collected and stored –70°C until standard DNA extraction with phenol/chloroform was performed (8).

Enrichment for (CA)_n Repeats Containing Fragments. The protocol followed in this paper has been described (16). The clone library was constructed with genomic DNA from the CL Brener clone (17). Genomic DNA was digested with *Mse*I or *Sau*3AI, and the fragments (500 ng) were ligated with 50 pmols of the corresponding adapters at 14°C overnight with 40 units of T4 DNA ligase in a 20 μl of reaction volume. An *Mse*I adapter was formed with *Mse*I-20 (TCTCCAGCTCTCAGGCAT) and *Mse*I-11 (TAAATGGGTTG), and *Sau*3AI adapter was formed with *Sau*3AI-24 (CGGGAATTCGCTCGTCGCCATG) and a *Sau*3AI-10 (GATTCGATGT). After the ligation, the 3’ ends of the fragments were repaired with 0.5 unit of *Taq* polymerase and 100 μM dNTPs for 10 min at 72°C. To capture the fragments containing (CA)_n repeats, 100 ng of the ligation product was mixed with 20 pmols of the biotinylated oligonucleotide CATGATCCGACCGCAT(CA)_{12} in 50 μl of TE buffer (10 mM Tris-HCl, pH 8.0/1 mM EDTA, pH 8.0) containing 2 μM of the MseI-20 or *Sau*3AI-24 to avoid hairpin formation. Mineral oil was added, and the mixture was heated to 95°C for 10 min and then annealed at 60°C for 1 min. Forty-five microliters were collected, added to 100 μg of the prewashed streptavidin-coated magnetic beads M-280 (Dynal, Oslo), and resuspended in 500 μl of TE/1 M NaCl. After a 30 min incubation with gentle rotation at room temperature, the beads were collected by magnet, washed three times with 3X SSC/0.5% SDS at 60°C for 15 min, once each with TE/1 M NaCl at room temperature, and resuspended in 50 μl of TE. To obtain double-stranded fragments, the captured DNA was PCR-amplified in 50 μl containing 67 mM Tris-HCl (pH 8.8),

© 1998 by The National Academy of Sciences 0027-8424

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.
Construction of Genomic Library Enriched for (CA)\textsubscript{n} Repeats. PCR products before and after the affinity capturing of (CA)\textsubscript{n} repeats were eletrophoresed in 1.5% agarose gel and then transferred to nylon filter Biodyne B (GIBCO/BRL). Approximately equal amounts of DNA (100 ng) were applied per lane. The filter was hybridized with alkaline phosphatase-conjugated (CA)\textsubscript{12} oligonucleotide and binding the complexes to streptavidin-coated magnetic beads. This resulted in a 10- to 15-fold enrichment of (CA)\textsubscript{n} containing fragments, as revealed by Southern blotting (Fig. 1B). A clone library of these fragments was then constructed, from which we randomly selected 44 recombinant colonies. After hybridization with the (CA)\textsubscript{n} probe, 39 colonies were positive and 25 of them were

### RESULTS

Preparation of an Enriched Library and Characterization of (CA)\textsubscript{n} Repeats from T. cruzi Genome. To obtain a library enriched in (CA)\textsubscript{n} repeats, we employed affinity capture of genomic restriction fragments by hybridizing them with a biotinylated (CA)\textsubscript{12} oligonucleotide and binding the complexes to streptavidin-coated magnetic beads. This resulted in a 10- to 15-fold enrichment of (CA)\textsubscript{n} containing fragments, as revealed by Southern blotting (Fig. 1B). A clone library of these fragments was then constructed, from which we randomly selected 44 recombinant colonies. After hybridization with the (CA)\textsubscript{n} probe, 39 colonies were positive and 25 of them were

Table 2 shows the primer sequences for each marker. After PCR, the amplified microsatellites were loaded on a 6% denaturing polyacrylamide gel and analyzed on ALF sequencer using the Fragment Manager software (Pharmacia). To determine the allele size, the samples were compared directly with band sizes from an allelic ladder prepared by amplification of an artificial mixture of DNA from 60 T. cruzi strains.

**Construction of a Phylogenetic Tree.** To make phylogenetic inferences, we assumed a stepwise mutation model for the microsatellites and used as a measure of genetic distance between any two strains the minimum number of mutational steps necessary to transform one into the other. The microsatellite multilocus genotypes were transformed into 0 and 1 characters by using FACTOR program from PHYLIP package version 3.57c (18). These data were collected and used next in the MIX program also from PHYLIP with which the evolutionary relationships were estimated based in the Wagner parsimony method. The significance of the branching in the Wagner network was done by bootstrapping (19). We used the program SEQBOOT of the PHYLIP package to perform 1,000 bootstraps. The number of times that a given branching was observed was used to evaluate the robustness of the tree.
randomly selected and sequenced. Seventeen colonies (68%) showed CA-repeat lengths greater than six, and from these, we chose 12 for which we designed specific primer pairs. A single set of PCR cycling parameters was chosen to facilitate the amplification of several markers simultaneously. We have been successful in amplifying 10 of the microsatellite markers, and eight of the markers that showed size polymorphism (Table 2) were studied further.

**Polymorphism of CA Repeats in T. cruzi.** To test the variability of these eight microsatellite markers, 30 different T. cruzi strains and clones were typed (Table 1). An important distinction to be made here is between strains and clones. Strains or isolates are obtained from triatomid insect vectors or mammalian hosts of T. cruzi, including humans. Patients in endemic areas may be infected by multiple contacts with different triatomids, and these triatomids, in turn, may feed on different infected individuals. This promiscuity propitiates the formation of multiclonal populations in hosts and vectors, leading to the isolation of correspondingly multiclonal strains when grown in culture. Molecular studies had confirmed that some T. cruzi strains were multiclonal (8, 20), but the general prevalence of multiclonality was not known.

To establish the ploidy of T. cruzi, we initially typed six different clones by PCR with primers for all microsatellite loci. All of the clones showed amplification of one or two fragments of different size (Fig. 2A). We interpreted this result as an indication that T. cruzi is diploid, the two-peak patterns indicating heterozygosity at the respective loci and the one-peak patterns denoting homozygosity. We next typed 24 different strains at all the eight loci. Again, we observed amplification of one- or two-fragment peaks for most strains. However, six strains (Gambá 3, 1009, 1023, 1030, 1017, and JM) consistently showed amplification of three or four major fragments of different size, implicating that these strains are multiclonal (Fig. 2B). It was interesting to observe that in the initial experiments, none of the strains showed more than two alleles isolated from chronic patients (data not shown). To further verify this point, we examined another 25 different strains isolated from chronic patients with one single microsatellite (MCLE01). Two strains (520/5 and JJS) showed four and three peaks respectively, showing that multiclonality also can be observed in chronic patients. However, when data from the 24 strains analyzed with all eight loci summarized, there seemed to be an important difference in the prevalence of multiclonality from strains obtained from the insects, wild animal, or acute patients (6/11) and chronic patients (0/13).

**Tests for Hardy–Weinberg Proportions and Linkage Disequilibrium.** Phenotypes for the eight microsatellite repeats were established for 20 of the T. cruzi strains or clones (excluded from this analysis were the strains with more than two bands, the clones from the Y strain—YP1, YP2, YP3, and the clone from the 182 strain), and genotypes were inferred. The observed heterozygosity of the microsatellite markers for this set of strains ranged from 0.20 to 0.80 (Table 2). Statistical tests showed highly significant departures from Hardy–Weinberg proportions ($\chi^2$ values ranged from 25.3 to 118.3; all

---

**Table 2. Microsatellites markers isolated from T. cruzi**

<table>
<thead>
<tr>
<th>Marker</th>
<th>Size range alleles, bp</th>
<th>Repeat sequence</th>
<th>PCR primers 5' → 3'</th>
<th>Number of alleles found</th>
<th>Observed heterozygosity</th>
<th>Expected heterozygosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCLE01</td>
<td>110–150</td>
<td>(CA)$_9$</td>
<td>1 CTGCCATGTTTGTCCCT 2 CTGTACATATCGGACAGT</td>
<td>11</td>
<td>0.60</td>
<td>0.84</td>
</tr>
<tr>
<td>MCLE08</td>
<td>124–130</td>
<td>(CA)$<em>2$(CA)$</em>{12}$</td>
<td>1 ATGGACAAACAATTGGGAG 2 TGGGTATGCAATAATGTGAT</td>
<td>4</td>
<td>0.20</td>
<td>0.71</td>
</tr>
<tr>
<td>SCLE10</td>
<td>237–291</td>
<td>(GT)$<em>2$(TG)$</em>{10}$</td>
<td>1 GATCCGCAATAGGAAAAC 2 GTGCATGTTCATGGCCTT</td>
<td>12</td>
<td>0.80</td>
<td>0.87</td>
</tr>
<tr>
<td>SCLE11</td>
<td>139–157</td>
<td>(AC)$_9$</td>
<td>1 ACGACAAAAAGCCATATT 2 GATGCTAATCTGCTAAGTA</td>
<td>9</td>
<td>0.60</td>
<td>0.83</td>
</tr>
<tr>
<td>MCLF10</td>
<td>182–194</td>
<td>(CA)$<em>2$(CA)$</em>{14}$</td>
<td>1 GGCTAGCGGATTTATTCCC 2 ATCCCGTACCTCAGTACAC</td>
<td>5</td>
<td>0.35</td>
<td>0.53</td>
</tr>
<tr>
<td>MCLG10</td>
<td>151–187</td>
<td>(CA)$_8$</td>
<td>1 AGGAGTCAAAATATAAGGGA 2 ACCTGTGAAAGCCATCTAC</td>
<td>7</td>
<td>0.20</td>
<td>0.69</td>
</tr>
<tr>
<td>MCL03</td>
<td>257–319</td>
<td>(CT)$_4$(GT)$<em>2$CT AT(GT)$</em>{15}$</td>
<td>1 GGAACGGAATGAAGGCA 2 TCGAGGAGTACGAC</td>
<td>13</td>
<td>0.65</td>
<td>0.91</td>
</tr>
<tr>
<td>MCL05</td>
<td>194–228</td>
<td>(TC)$_9$(GT)$_4$</td>
<td>1 TTAACAGCCTCTATGCTCTC 2 CCTGAGCAGTACAAGGAC</td>
<td>12</td>
<td>0.55</td>
<td>0.90</td>
</tr>
</tbody>
</table>
were significant at $P < 0.001$). Indeed, as can be seen in Table 2, expected heterozygosities ($H = 1-\sum x_i^2$, where $x_i$ is the frequency of the $i^{th}$ allele) calculated assuming Hardy–Weinberg were very different from the observed ones. Statistical tests for all of the 28 pairwise comparisons possible for all of the eight microsatellite loci showed that all of these loci are associated in very strong linkage disequilibrium ($P < 0.001$).

**Phylogenetic Relationships.** *T. cruzi* multilocus genotypes were constructed based on the eight microsatellites repeats studied. Each strain produced a highly individual genotype. We utilized these genotypes to investigate the phylogenetic relationships and *T. cruzi* population structure by using the maximal parsimony approach. It is believed that microsatellites evolve by DNA strand slippage, with most mutations involving the gain or loss of only a single repeat unit (21). Considering each microsatellite allele as one state from a multistate character and assuming that the most frequent allele was the ancestral state, we can establish a genetic distance measure between any two strains as being the number of mutational steps necessary to transform one into the other. We then used the Wagner parsimony algorithm from the MIX computer program of the PHYLP package phylogenetic tree (18) to obtain one most parsimonious unrooted Wagner network that is shown in Fig. 3, where the figures in parenthesis indicate the number of times that the branch was observed in 1,000 bootstraps. The minimal distance between two *T. cruzi* strains was 17 mutational events, and the overall dispersion pattern based on microsatellite genotypes shows that the *T. cruzi* strains are very distant from each other. This is especially significant because most strains were from the same geographic area. The only significant clusters observed were one involving the 115, 226, and 231 strains, another with 167, 182, and CL Brener, and a third one with 1001 and Colombiana. These same strains also formed clusters when studied with other approaches used to characterize genetic variability of anonymous DNA sequences in *T. cruzi* strains (11). On the other hand, we could observe a clear separation of strains based on rRNA typing (14). Strains classified as belonging to group 1 were found at one extremity of the network whereas strains belonging to groups 1/2 and 2 were found at the other end (Fig. 3). This clear separation was observed in 100% of the bootstraps.

**DISCUSSION**

Our studies with CA-repeat microsatellites provide several insights in the genetic structure of *T. cruzi* populations. First, the fact that all clones showed patterns of one or two microsatellite bands with different loci gives strong support to the notion that *T. cruzi* is diploid in most, if not all, of its genome. Most of the previous information supporting the diploidy was derived from isoenzyme studies or on hybridization on restricted DNA or chromosomal bands separated by pulse-field gel electrophoresis (5, 12, 22, 23).

Second, our microsatellite studies showed drastic departures from Hardy–Weinberg expectations and strong linkage disequilibrium, suggesting that sexual reproduction is rare or absent in *T. cruzi* and that the population structure is clonal, as had been suggested on the basis of isoenzyme studies (5, 6, 7). Thus, each clone represents a lineage that reproduces by binary division and remains unaltered for a large number of generations until mutations occur.

Third, we observed several strains, especially those obtained from insects, wild animals, and acute patients with more than two bands in a single microsatellite. This observation is best compatible with a multiclonal composition of these strains.
Previously, a few studies already had demonstrated that clones obtained from some strains could display different restriction fragment length polymorphism patterns of kinetoplast DNA minicircles (20) or were heterogeneous in DNA fingerprinting (8). However, our microsatellite studies provide the first direct demonstration of frequent multiclonality of strains, which indeed was to be expected on theoretical grounds. We also have to consider that we are probably only observing the tip of the multiclonality iceberg because growth in culture may entail the loss of some clones incapable of adapting to the new in vitro environment. Actually, we also should expect that the infection of the human host also might entail reduction of multiclonality. *T. cruzi* has existed as a species for ten or hundreds of millions of years but has only met the human host after the colonization of the American continent 10,000–30,000 years ago (24). Thus, it is quite conceivable that some clonal lineages found in wild animals and insect hosts are incapable of infecting humans. Wild strains of *T. cruzi* presumably represent swarms of clones that may present symbiotic relationships but also certainly compete fiercely for available resources. The patterns of interactions within the swarm may change drastically in different environments, especially in the passage of the clonal population from the invertebrate vector to man. One would expect that some clones would be eliminated by their inability to compete and propagate in the human environment or by the action of the host defense mechanisms, decreasing the complexity of the infecting strain. Thus, it is not surprising that we found that the prevalence of multiclonality was much higher in strains obtained from insects, wild animals, or patients with acute disease than from chronic patients. Studies with microsatellites also provide a very useful means to test whether a given strain contains one or several clones. If a strain consistently displays one or two bands on amplification with a battery of several microsatellites, as we have seen for several of the strains that we studied, monoclonality may be safely assumed. This is particularly relevant in some experimental studies because occasionally some strains with interesting biological properties are incapable of being cloned in culture. Furthermore, these microsatellites markers may have a major impact in the understanding of the pathogenesis of the Chagas’ disease through the genetic typing of *T. cruzi* present directly in the host tissue.

Fourth, with the polymorphic CA-repeat microsatellites, we found a remarkable genetic diversity in *T. cruzi*. Furthermore, no other previous method of genetic analysis illustrated so cogently this genetic individuality in *T. cruzi* because no single repeated multilocus genotype was detected among the strains analyzed. Indeed, with the study of 15 isoenzymes, identical phenotypes were observed in geographically unrelated clones (5). When we used parsimony to display the genetic relationship among the clonal lineages in an unrooted Wagner network (Fig. 3), we did not observe any geographical clustering of the clones and strains and the minimal genetic distance separating them was 17 mutational steps. However, when we typed the clones and strains according to a ribosomal DNA restriction fragment length polymorphism (14), there was an excellent correlation of types 1 and 2 with the tree topography (Fig. 3). The only two strains that belonged to 1/2 (115 ± 226) clustered with group 2, suggesting that conceivably the latter may be the ancestral state.

In summary, we show in this article that microsatellite polymorphisms typed by PCR amplification represent a powerful tool for the understanding of the population structure and dynamics of *T. cruzi*. The simple and fast strategy used here is applicable to broad range of parasites permitting the discovery of CA repeats and other microsatellites. This strategy will be fundamental for physical mapping in the parasites included in the Parasite Genome Initiative supported by the World Health Organization (25). We have provided evidence that *T. cruzi* is diploid, that it has a genetically extremely diverse clonal population structure, and that clones associate to form multiclonal strains. Indeed, microsatellite typing is a direct and useful means of testing the monoclonality of a strain. Our next goal will be to achieve the genetic typing of single parasites isolated by cell sorters, dispensing cloning, and permitting considerable refinement of *T. cruzi* population studies.

This work was supported by Conselho Nacional de Pesquisas (Brazil), Comissão de Aperfeiçoamento de Pessoal de Nível Superior, and Programa de Núcleos de Excelência. We thank Prof. Egler Chiari from the Department Parasitology of the Universidade Federal de Minas Gerais for providing the strains and clones of *T. cruzi* used in this study. We are grateful to Anna Izabel R. Melo for helping in the microsatellite typing.