Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*

Carlos A. Machado* and Francisco J. Ayala

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697

Simple phylogenetic tests were applied to a large data set of nucleotide sequences from two nuclear genes and a region of the mitochondrial genome of *Trypanosoma cruzi*, the agent of Chagas' disease. Incongruent gene genealogies manifest genetic exchange among distantly related lineages of *T. cruzi*. Two widely distributed isoenzyme types of *T. cruzi* are hybrids, their genetic composition being the likely result of genetic exchange between two distantly related lineages. The data show that the reference strain for the *T. cruzi* genome project (CL Brener) is a hybrid. Well-supported gene genealogies show that mitochondrial and nuclear gene sequences from *T. cruzi* cluster, respectively, in three or four distinct clades that do not fully correspond to the two previously defined major lineages of *T. cruzi*. There is clear genetic differentiation among the major groups of sequences, but genetic diversity within each major group is low. We estimate that the major extant lineages of *T. cruzi* have diverged during the Miocene or early Pliocene (3–16 million years ago).

The protozoan *Trypanosoma cruzi* is very heterogeneous at the phenotypic and genetic levels. Strains of *T. cruzi* show large differences in growth rate, histotropism, antigenicity, pathogenicity, infectivity of potential insect vectors, drug susceptibility, chromosome number, and DNA content. The first survey of the genetic diversity of this parasite based on electrophoretic data from six enzymes showed that there were at least three main strain groups or zymodemes (Z1, Z2, Z3) (1). More comprehensive multilocus enzyme electrophoresis (MLEE) studies defined 43 isoenzyme types on the basis of allozyme profiles from 15 loci (2, 3). Further work based on data from MLEE, random amplified polymorphic DNA (RAPDs), and polymorphism of rDNA and miniexon genes (4–6) suggested the presence of two major lineages in *T. cruzi*, one corresponding to Z1 (isoenzyme types 1–25) and the other to Z2 and Z3 (isoenzyme types 26–43) (7). Recently, an effort was made to unify the different classifications of *T. cruzi* (8), leading to the definition of "*T. cruzi* I" and "*T. cruzi* II" (equivalent to Z1 and Z2). The taxon designation of Z3 was left to be decided by further studies. Additional RAPD markers (9, 10) indicated that *T. cruzi* II should further be subdivided into five distinct lineages (IIa–IIe).

Analyses of MLEE, RAPD, and microsatellite data (2, 3, 5, 11, 12) suggest that the population structure of *T. cruzi* is predominantly clonal, with sexuality having no or a very limited influence on the genetic structure and evolution of the parasite (2, 3). However, a few MLEE, RAPD, and restriction fragment length polymorphism studies (11, 13–16) have suggested that the genetic diversity of *T. cruzi* is caused not only by clonal divergence but also by genetic exchange. But the problems of homoplasy, lack of resolution, and uncertain homology associated with MLEE, RAPDs, and microsatellites hamper the use of the available data to evaluate the extent of genetic exchange in *T. cruzi*. Those problems are minimized by using DNA sequence data, which, moreover, can be analyzed with the powerful analytical tools of molecular population genetics and molecular phylogenetics.

DNA sequence data collected from multiple unlinked loci allow one to rigorously test predictions about expected patterns of phylogenetic divergence in genes collected from organisms that propagate clonally. Under strict clonality, all neutrally evolving genes from a lineage should share the same history of divergence, because the complete genome is transmitted intact each generation (17–20), whereas genes from sexually reproducing organisms tend to have very different histories, because the genome is recombined each generation. Therefore, incongruent gene genealogies imply genetic exchange, if one can reject the null hypothesis that the genealogies are not significantly different.

Additionally, when reproduction is attained only by mitotic division or parthenogenesis, long-term asexual reproduction in diploid taxa should lead to complete heterozygosity (21–23). Therefore, high levels of divergence between the two copies of a gene in every individual should be observed (24, 25), a phenomenon termed intracellular allele sequence divergence (ASD) (24). Because rare sexual reproduction can erase any evidence of ASD (24), it is predicted that if *T. cruzi* is composed of strictly asexual diploid lineages, nucleotide sequences should provide evidence of ASD. However, lack of significant ASD in asexual diploid taxa may be a consequence of automixis (21, 22), gene conversion, or mitotic recombination events (24).

We test these predictions with nucleotide sequence data from two single-copy protein-coding nuclear genes and from a region of the mitochondrial genome of *T. cruzi*, obtained from a large sample of strains that represent much of the known genetic diversity of the parasite.

Materials and Methods

Samples. Forty-six strains representing 12 isoenzyme types of *T. cruzi* were used (Table 1). Each strain name in the text and figures has a prefix number that corresponds to its isoenzyme type (as designed in ref. 3).

Molecular Methods. Nucleotide sequences were obtained from two single-copy nuclear genes that encode the enzymes trypanothione reductase (TR) (26) and dihydrofolate reductase–thymidylate synthase (DHFR-TS) (27) and from a region of the mitochondrial genome partially encompassing the maxicircle-encoded genes cytochrome oxidase subunit II (COII) (28) and NADH dehydrogenase subunit 1 (ND1). Sequences of 1,290, 1,473, and 1,226 bp were obtained, respectively, for TR, DHFR-TS, and COII-ND1. The following primers (5′-3′) were used for PCR: for the COII-ND1

Table 1. The 46 *T. cruzi* strains

Zymodeme	Lineage	Isoenzyme type	Strains
Z1	*T. cruzi* I	12	TEH cl2, cl92, CEPA EP*, Vin C6, FLORIDA C16
		17	X10 cl1, SABP3, A80, A92*, MA-V*
		19	OPS21 cl11, CUTIA cl1, 133 79 cl7, V121*, 26 79
		20	CUICA cl1, SO34 cl4, P209 cl1, 85/818, P0AC*, Esquilo cl1
		?[†]	SC13
Z2	*T. cruzi* IIb	30	ESMERALDO cl3, X-300*
		32	TU18 cl2, CBB cl3, MSC2, MCV*, MVB cl8*
	T. cruzi IIc	35	M6241 cl6
		36	M5631 cl5[‡], CM 17, X110/8, X9/3*, X109/2*
	T. cruzi IId	39	SO3 cl5, EPP, PSC-O, 86-1*
	T. cruzi IIe	43	CL F11F5[§], TULAHUEN cl2, P63 cl1*, 86/2036*, P251*, VMV4*
Z3	*T. cruzi* IIa	27	CANIII cl1
		nd(27)[†]	EP 255[¶]

Zymodeme and isoenzyme type as in refs. 1 and 3, respectively; lineages, refs. 8 and 10. Sequences were also obtained from the bat trypanosomes *T. cruzi marinkellei* 593 (B3) and *T. vespertilionis* (N6).
*Only COII-ND1 sequenced.
[†]The MLEE profiles of SC13 and EP 255 do not fit any isoenzyme type of ref. 3, but the profile of EP 255 is close to type 27 [nd(27)].
[‡]COII-ND1 could not be fully amplified.
[§]Clone CL Brener, the reference strain for the *T. cruzi* genome project.
[¶]The TR sequence could not be obtained.

region, ND1.3A (GCTACTARTTCACTTTCACATTC), COII.2A (GCATAAATCCATGTAAGACMCCACA); for the DHFR-TS gene, DH1S (CGCTGTTTAAGATCCGNATGCC), DH3A (CGCATAGTCAATGACCTCCATGTC); for the TR gene, TRY2S (ACTGGAGGCTGCTTGGAACGC), TRY2A (GGATGCACACCRATRGTGTTGT); where A and S stand, respectively, for antisense and sense DNA strands. PCR conditions to amplify the three markers were: 30-sec denaturation at 94°C, primer annealing for 1 min at 58°C, and primer extension for 2 min at 72°C, for a total of 30 cycles. PCR products were purified with the Wizard PCR Preps DNA Purification Kit (Promega). The PCR primers were used for bidirectional sequencing, and internal sequencing primers were designed when partial sequences were obtained; COII-ND1 region: COII.A400 (CTCCTATTACAAC-CAATAAACATC); DHFR-TS gene: DHSEQS (AGCATTGR-GACRGTCGTCTACTG), DHSEQA (ACCCTGTCCGTCAT-AGTTG); TR gene: TRYSEQS (CGAATGARGCATT-YTACCTG), TRYSEQA (TACTCGTCCACCTGCACACCAC). PCR products were sequenced in both directions in an ABI 377 automatic sequencer (Applied Biosystems) by using standard protocols. The use of dRhodamine chemistry made possible to detect polymorphic sites as unambiguous double peaks in the chromatograms of the directly sequenced PCR fragments from the nuclear genes. If the sequence from a strain was polymorphic, the PCR product was cloned by using the TA-cloning kit (Invitrogen). Ten to fifteen cloned PCR fragments from each strain were sequenced to identify the different haplotypes. That strategy was used to detect evidence of ASD.
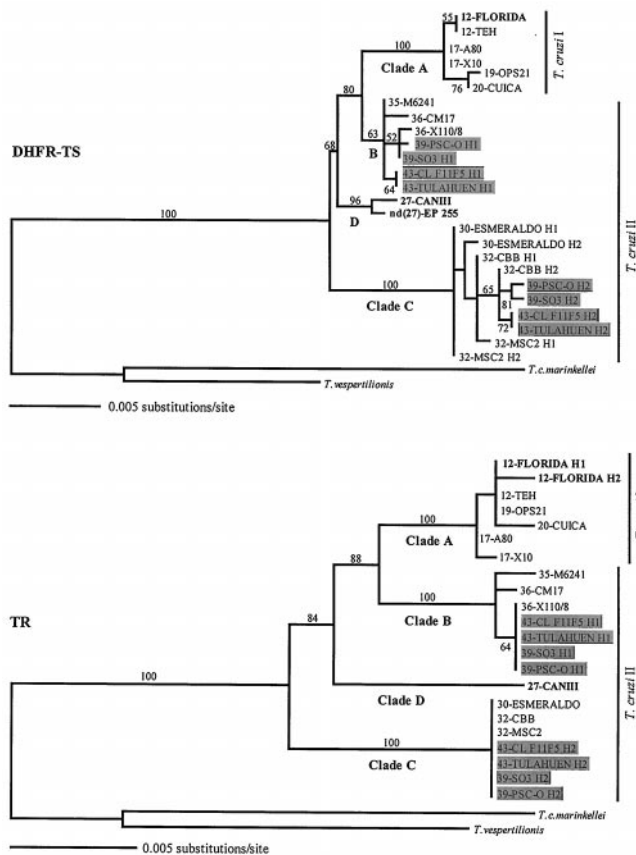
Phylogenetic Analyses. We used PAUP* (29), with sequences from the bat trypanosomes *T. cruzi marinkellei* [593 (B3)] and *Trypanosoma vespertilionis* (N6) as outgroups. Initially, a Neighbor-Joining (NJ) tree (30) was reconstructed for each data set by using Tamura–Nei distances (31). Each topology was used to estimate maximum likelihood parameters for different models of nucleotide substitution, and the most appropriate model for each data set was chosen by using a likelihood ratio test (32). Maximum likelihood
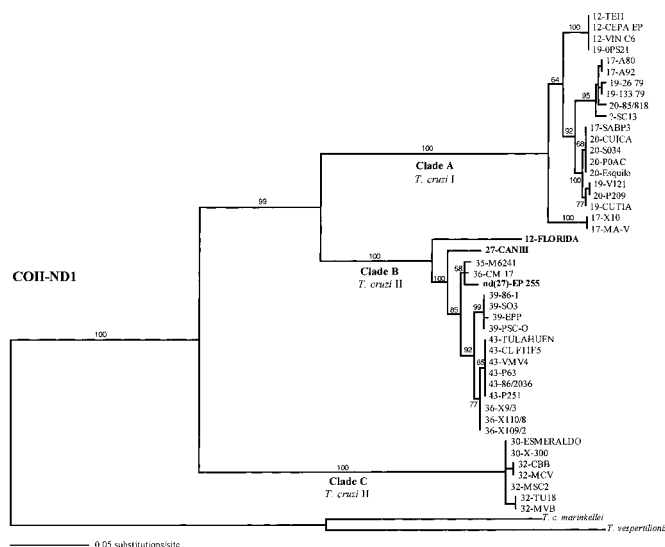
Fig. 1. Genealogical relationships among COII-ND1 sequences from 45 *T. cruzi* strains. The prefix number of each strain is the isoenzyme type of ref. 3. ML analyses use the general reversible model with rate heterogeneity (39) (REV+Γ; −ln (L) = 3601.76615; α = 1.95234). The topology of the NJ and parsimony trees [412 steps, consistency index (C.I.) = 0.794; ingroup only: 250 steps, C.I. = 0.856] are almost identical. Numbers at the branches are bootstrap values >50% (500 replications) for the same nodes of the NJ tree. The three taxa with positions incongruent with the nuclear genealogies (Fig. 2) are in bold. Main clades are named with an uppercase letter. Strain 12-FLORIDA is classified as *T. cruzi* I, but its mitochondrial sequence clusters with strains classified as *T. cruzi* II.

(ML) trees for each data set were then reconstructed by using the parameters of the models estimated from the data and a tree-bisection-reconnection branch-swapping algorithm. Strict and relaxed topological congruency tests were conducted by using the Shimodaira–Hasegawa test (33). The strict congruency test is based on the assumption that the same genealogy applies to the three sets of gene sequences. By using a given data set and its set of ML parameters, the likelihood of the ML tree is compared with the likelihood obtained by forcing the data to fit the topology of the ML trees obtained for the other data sets. The common assumption for the relaxed congruency tests is that the gene trees share basic topological similarities defining the major clades. These tests compare the likelihoods of the ML tree and the best congruent tree constrained by the basic topological similarities. The best congruent tree is reconstructed by a heuristic search constrained on an unresolved topology consistent with a given phylogenetic hypothesis (e.g., a topology in which only the position of major incongruent strains is constrained, or in which strict congruency is forced only for strains of one major clade).

Estimating Divergence Times. Available sequences for COII-ND1 and TR from the African parasite *Trypanosoma brucei* (GenBank accession nos. M94286 and X63188) were used to calibrate rates of substitution and estimate the age of the *T. cruzi* clade. No DHFR-TS sequence is available for *T. brucei*. Substitution rates were calibrated by using 100 million years (Myr) for the divergence between *T. cruzi* and *T. brucei* (34, 35). Rate constancy across all branches of the gene genealogy was tested by using a likelihood ratio test (36). Times of divergence were estimated by multiplying the ML branch lengths, estimated under the constraint of a molecular clock, by the ratio of divergence time to branch length of the reference node (the most recent common ancestor of *T. brucei* and *T. cruzi*). Confidence intervals were defined as plus or minus twice the standard error of the branch length times the rate of substitu-

DHFR-TS

```
                            55 ┌ 12-FLORIDA
                               ├ 12-TEH
                      100      │ 17-A80
                  ┌──────┐  17-X10
          Clade A │          ├ 19-OPS21
       80 │                76 └ 20-CUICA
    ┌─────┤           35-M6241
    │     │           36-CM17
    │     │        63┌ 36-X110/8
    │   68│        52├ 39-PSC-O H1
    │     │       B  ├ 39-SO3 H1
    │     │          ├ 43-CL F11F5 H1
    │     │        64└ 43-TULAHUEN H1
    │     │     96┌ 27-CANIII
    │     └──────┤ nd(27)-EP 255
    │          D
100 │              30-ESMERALDO H1
    │              30-ESMERALDO H2
    │              32-CBB H1
    │              32-CBB H2
    │           65┌ 39-PSC-O H2
    └──────┐      ├ 39-SO3 H2
     Clade C    81├ 43-CL F11F5 H2
            100 72└ 43-TULAHUEN H2
                   32-MSC2 H1
                   32-MSC2 H2
```

T. cruzi I / *T. cruzi* II / *T.c.marinkellei*

T. vespertilionis

—— 0.005 substitutions/site

TR

```
                            ┌ 12-FLORIDA H1
                            ├ 12-FLORIDA H2
                    100     │ 12-TEH
                 ┌──────┐   ├ 19-OPS21
         Clade A │          ├ 20-CUICA
      88 │                  │ 17-A80
   ┌─────┤                  └ 17-X10
   │     │           35-M6241
   │     │     100   36-CM17
   │     └──────┐    36-X110/8
   │       Clade B ┌ 43-CL F11F5 H1
   │              64├ 43-TULAHUEN H1
   │                ├ 39-SO3 H1
84 │                └ 39-PSC-O H1
   │                   27-CANIII
   │         Clade D
   │                   30-ESMERALDO
   │                   32-CBB
100│              100   32-MSC2
   │           ┌────── 43-CL F11F5 H2
   └───────────┤  43-TULAHUEN H2
         Clade C   39-SO3 H2
                   39-PSC-O H2
```

T. cruzi I / *T. cruzi* II / *T.c.marinkellei*

T. vespertilionis

—— 0.005 substitutions/site

Fig. 2. Genealogical relationships among *T. cruzi* strains based on ML analyses of sequences from the nuclear genes DHFR-TS [$-\ln(L) = 2836.17289$; $\alpha = 11.57032$; ts/tv ratio $= 4.21936$] and TR [$-\ln(L) = 2464.72229$; $\alpha = \infty$; ts/tv ratio $= 1.99505$]. The Hasegawa–Kishino–Yano model with rate heterogeneity (HKY$+\Gamma$) is used for the analyses (40). Only a subset of the sequenced taxa is shown to help in visualizing the incongruences with the mitochondrial genealogy. Analyses with all taxa yield almost identical trees. Numbers at the branches are bootstrap values >50% (500 replications) for the same nodes of the NJ tree. The parsimony and ML topologies are almost identical [DHFR-TS: 125 steps, consistency index (C.I.) = 0.936; TR: 102 steps, C.I. = 0.941]. The three taxa with incongruent positions in the mitochondrial genealogy (Fig. 1) are in bold. The hybrid strains (isoenzyme types 39 and 43) are shown in shaded boxes.

tion. Branch lengths and standard errors were obtained by using the program PAML ver. 2.0a (37).

Results

Gene Genealogies. A test that looks at the decay of significant levels of linkage disequilibrium with distance (38) does not provide evidence of intragenic recombination for any of the three data sets (not shown). Therefore, the history of the sequences from each gene can be represented by a gene genealogy (phylogenetic tree). Genealogies of all of the mitochondrial sequences and a subset of the nuclear sequences, are shown in Figs. 1 and 2, respectively. The topologies of the genealogies do not fully match previous phenograms reconstructed with MLEE and RAPD data that suggested the presence of two major phylogenetic lineages in *T. cruzi* (5, 6, 11). Instead, the sequences cluster in three or four distinct clades (Figs. 1 and 2). One clade includes all sequences from strains of zymodeme Z1 (represented by strains from isoenzyme types 12, 17, 19, and 20, hereafter referred to as clade A). However, sequences from strains of Z2 and Z3 (isoenzyme types 26–43) are not monophyletic and cluster in two (mitochondrial data) or three (nuclear data) distinct clades (hereafter clades B, C, and D). The most important difference from previous phenograms is that all sequences from

Table 2. Average uncorrected divergence per base pair among sequences from each major group of *T. cruzi* sequences

Comparison	COII-ND1*	DHFR-TS	TR[†]
Within clade A	0.01037	0.00084	0.00147
Within clade B	0.00371	0.00129	0.00155
Within clade C	0.00098	0.00197	0.00000
Clade A vs. D	—[‡]	0.00963 (12)	0.01374 (15)
Clade A vs. B	0.07297 (69)	0.00735 (9)	0.01021 (10)
Clade A vs. C	0.09533 (99)	0.01396 (15)	0.01615 (19)
Clade D vs. B	—[‡]	0.00655 (7)	0.01460 (18)
Clade D vs. C	—[‡]	0.01052 (10)	0.01628 (21)
Clade B vs. C	0.09041 (104)	0.01077 (10)	0.01731 (21)
Clade A vs. *T. c. marinkellei*	0.12262 (132)	0.04926 (71)	0.03786 (47)
Clade A vs. *T. brucei*	0.24448 (282)	—[§]	0.26546 (341)

The number of fixed nucleotide differences is shown in parentheses. Clades A–D as in Figs. 1 and 2. Comparisons with the outgroup are shown only for clade A. Distances estimated using the program SITES (41).
*The COII-ND1 sequences from strains 12-FLORIDA, 27-CANIII, and nd(27)-EP255 are not included in the calculations.
[†]Because the TR sequence from strain nd(27)-EP255 could not be collected, clade D is defined by a single sequence (27-CANIII).
[‡]There is no clade D in the mitochondrial genealogy.
[§]No DHFR-TS sequence is available for *T. brucei*.

strains classified as isoenzyme types 30 and 32 belong to the most divergent clade of *T. cruzi* sequences (clade C) and are not part of the clade formed by the other strains of Z2 and Z3. Genetic distances and fixed nucleotide differences among the major clades are in Table 2. There is clear divergence (albeit low in the nuclear genes) among the major groups of *T. cruzi* but very low divergence within each group.

Genetic Exchange. The most striking evidence of genetic exchange is the inference that the nuclear genome of isoenzyme types 39 and 43 is hybrid. None of the sampled strains is heteroplasmic for the mitochondrial sequences. However, several strains are heterozygotic for the nuclear sequences, and the haplotypes from all strains of isoenzyme types 39 and 43 are fairly divergent (Table 3); each haplotype is either identical or shows few differences with the sequences from types 35 and 36 or 30 and 32 (Table 3 and Fig. 2). Laboratory contamination is ruled out, because each strain from types 39 and 43 had a unique mitochondrial sequence distinct from that of any other isoenzyme type, and because other strains from types 39 and 43 that were independently isolated also had polymorphic nuclear sequences. This haplotype pattern is presumably the result of one or more hybridization events between strains from clades B and C, with no detected introgression of mitochondrial haplotypes from clade C into the hybrid genotypes (Fig. 1). Our results confirm previous suspicions about the hybrid nature of isoenzyme types 39 and 43 on the basis of MLEE and RAPD data (7, 9, 11) and confirm that the reference strain for the *T. cruzi* genome project (43-CL F11F5 or CL Brener) (42) is a hybrid.

Interestingly, although the origin of the hybrid genotypes is explained by the occurrence of genetic exchange, their maintenance as stable hybrid genotypes over time and space (most of the isolates have been collected during the last 40 years in different geographic regions) is consistent with the prevalence of asexual reproduction in *T. cruzi*. Automixis or Mendelian sexuality in the hybrids should generate strains homozygous for alleles from clades B or C that should also carry the distinct mitochondrial haplotype from isoenzyme type 39 or 43. That is not observed in our sample. However, the sequences from strains 36-X9/3, 36-X110/8 and 36-X109/2, which are homozygous for the nuclear genes, are more closely related (COII-ND1) or identical (TR) to sequences from types 43 and 39 than to sequences from other strains of type 36 (Table 3, Figs. 1 and 2). This could result from one or several instances of automixis or

Table 3. Variable sites in the DHFR-TS and TR sequences from strains of isoenzyme types 30, 32, 35, 36, 39, and 43

DHFR-TS

Strain/Site	153	163	222	288	294	324	334	408	410	446	462	522	568	649	724	986	997	1045	1095	1128	1158	1246	1296	1356	1416
35-M6241 cl6	C	G	C	T	C	G	C	G	A	A	T	C	T	C	G	C	A	C	G	A	C	T	A	G	G
36-M5631 cl5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.
36-CM 17	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
36-X110/8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	T	.
39-PSC-O H1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	T	.
39-SO3 cl5 H1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.
43-CL F11F5 H1	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
43-Tulahuen cl2 H1	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
30-Esmeraldo cl3 H1	T	T	T	C	.	.	.	.	G	.	C	T	C	.	.	.	G	.	T	G	A	G	G	.	.
30-Esmeraldo cl3 H2	T	T	T	C	.	.	.	.	G	.	C	T	C	T	.	T	G	.	T	G	A	G	.	.	.
32-MSC2 H1	T	T	T	C	.	.	.	.	G	.	C	.	C	.	.	T	G	.	T	G	A	G	G	.	A
32-MSC2 H2	T	T	T	C	.	.	.	.	G	.	C	T	C	.	.	.	G	.	T	G	A	G	G	.	.
32-TU18 cl2 H1	T	T	T	C	.	.	.	.	G	.	C	.	C	.	.	T	G	.	T	G	A	G	G	.	.
32-TU18 cl2 H2	T	T	T	C	T	C	.	.	G	.	C	.	C	.	.	T	G	.	T	G	A	G	G	.	.
39-SO3 cl5 H2	T	T	.	C	T	C	.	.	G	.	C	.	C	.	.	T	G	T	T	G	A	G	G	.	.
39-PSC-O H2	T	.	.	C	T	C	.	.	G	G	C	.	C	.	.	T	G	.	T	G	A	G	G	.	.
43-CL F11F5 H2	T	T	T	C	T	C	.	.	G	.	.	.	C	.	.	T	G	.	T	G	A	G	G	.	.
43-Tulahuen cl2 H2	T	T	T	C	T	C	.	.	G	.	.	.	C	.	.	T	G	.	T	G	A	G	G	.	.
Amino acid change		V						Q	E																
		L						R	G																

TR

Strain/site	198	222	231	270	372	466	486	549	594	612	615	649	736	756	813	876	903	1008	1011	1045	1058	1206	1207	1251	1272
35-M6241 cl6	G	G	T	A	C	C	G	C	C	T	A	G	G	G	T	C	T	T	G	A	C	A	G	A	T
36-M5631 cl5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
36-CM 17	A	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
36-X110/8	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
39-PSC-O H1	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
39-SO3 cl5 H1	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
43-CL F11F5 H1	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
43-Tulahuen cl2 H1	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
30-Esmeraldo cl3	A	.	C	G	T	A	T	T	T	C	C	A	.	A	C	A	C	C	T	G	A	C	A	C	G
32-TU18 cl2	A	.	C	G	T	A	T	T	T	C	C	A	.	A	C	A	C	C	T	G	A	C	A	C	G
39-SO3 cl5 H2	A	.	C	G	T	A	T	T	T	C	C	A	.	A	C	A	C	C	T	G	A	C	A	C	G
39-PSC-O H2	A	.	C	G	T	A	T	T	T	C	C	A	.	A	C	A	C	C	T	G	A	C	A	C	G
43-CL F11F5 H2	A	.	C	G	T	A	T	T	T	C	C	A	.	A	C	A	C	C	T	G	A	C	A	C	G
43-Tulahuen cl2 H2	A	.	C	G	T	A	T	T	T	C	C	A	.	A	C	A	C	C	T	G	A	C	A	C	G
Amino acid change					H								G									I	T	K	V
					N								S									V	N	N	I

Numbers correspond to nucleotide sites in the GenBank reference sequences. The suffixes H1 and H2 for isoenzyme types 39 and 43 refer to the arbitrarily defined haplotypes 1 and 2, which are homologous, respectively, to the sequences from isoenzyme types 35–36 and 30–32. The two DHFR-TS haplotypes from 32-CBB and 32-TU18 are identical. The TR sequences from 32-CBB and 32-MSC2 are identical to those from 30-Esmeraldo cl3 and 32-TU18. The sequences from 39-EPP and 39-PSC-O are identical.

Mendelian sexuality in a putative hybrid ancestor. Alternatively, it could be interpreted as evidence of two independent hybridization events generating isoenzyme types 39 and 43.

The second evidence of genetic exchange comes from the observation that the three gene genealogies are incongruent (Table 4), which is not expected under strict clonality. Forcing each data set to fit the ML topologies estimated for the other two data sets generates trees that are significantly worse than the ML trees (Table 4). Therefore, the incongruence is not due to random or systematic errors in estimating the genealogies. A higher rate of substitution in the genes from some *T. cruzi* lineages is also ruled out, because likelihood ratio tests do not reject the molecular clock hypothesis for the three data sets (not shown). The incongruence is more striking in the comparisons between the genealogy of the mitochondrial region and the genealogies of the two nuclear genes (Figs. 1 and 2), where three conspicuous cases of mitochondrial introgression are observed [strains 12-FLORIDA, 27-CANIII, and nd(27)-EP255]. These incongruences are most likely the result of three separate instances of genetic exchange among distantly related lineages of *T. cruzi*. In each case, there has been introgression of mitochondrial haplotypes between strains from divergent clades, and in all cases the source of the introgressed haplotype is a group of lineages that appear basal to clade B. Congruency is also rejected if it is constrained only for the three incongruent strains (Table 4). However, the large differences in log likelihood between those unconstrained topologies and the strictly congruent trees (Table 4) suggest several cases of mitochondrial introgression among closely related strains, which is also suggested by the observation that several strains from the same isoenzyme types carry fairly distinct mitochondrial haplotypes (Fig. 1).

The two nuclear genealogies show the same four major clades but are incongruent (Table 4). This is seemingly due to genetic exchange among closely related lineages, given that the sequences of the two nuclear genes from each strain fall in the same major clade. When topological congruency is assessed separately for each major clade, the DHFR-TS data reject only congruency for clade C, whereas the TR data reject only congruency for clade B (Table 4). These results suggest the occurrence of genetic exchange among

Table 4. The tests of congruency among the nuclear and mitochondrial genealogies

Tree topology or phylogenetic hypothesis	COII-ND1		DHFR-TS		TR	
	-ln (L)	P	-ln (L)	P	-ln (L)	P
Strict congruency test						
COII-ND1 ML topology	**3592.25027**		3015.04197	0.0011	2561.47821	0.0140
DHFR-TS ML topology	4013.97303	<0.0001	**2841.36899**		2593.27602	0.0324
TR ML topology	3875.89194	<0.0001	2956.60300	0.0076	**2503.77611**	
Relaxed congruency tests						
Congruent to basic nuclear topology	3742.62997	<0.0001				
Congruent to basic mitochondrial topology			2938.09706	0.0079	2540.51230	0.0082
Clade A congruent to TR or DHFR-TS ML topology			2847.54852	0.7799	2509.74392	0.5044
Clade B congruent to TR or DHFR-TS ML topology			2850.76382	0.7263	2522.05809	0.0191
Clade C congruent to TR or DHFR-TS ML topology			2918.74561	0.0169	2503.77611*	1.0000

P is the probability of getting at least the observed Δ -ln (L) value under the null hypothesis that there is no significant difference between the ML tree and the alternative topology (SH test; RELL approximation). ML values are in boldface. Analyses are only for strains with all three genes sequenced. The DHFR-TS topology compared with the TR data only includes the H2 haplotypes from clade C (results with the H1 haplotypes are similar). The first two relaxed congruency tests compare the likelihood of the ML tree for the given mitochondrial or nuclear dataset with that of a tree in which only the position of strains 12-FLORIDA, 27-CANIII, and nd(27)-EP 255 is forced to be congruent with their position in the nuclear or mitochondrial topologies, respectively. The last three rows are for relaxed congruency tests in which strict congruency is forced for one major clade at a time.
*Identical to the ML value, because all TR sequences from clade C are identical.

closely related strains from these two clades. The incongruence among strains of clade C by the DHFR-TS data is due to the polytomy of clade C in the TR tree (all sequences are identical), which could be the result of recent genetic exchange among those strains or simply a reflection of lower genetic diversity in the TR gene. However, this last explanation is unlikely, given that TR is more polymorphic than DHFR-TS (Table 2).

Lack of ASD. ASD is observed in the nuclear genes of the hybrid isoenzyme types, but it is the result of a recent hybridization event and not of ancient asexuality. No signs of ASD are found in any of the other strains. Most are homozygous for TR and DHFR-TS despite the observed divergence among sequences from distantly related strains (Tables 2 and 3). The lack of ASD in *T. cruzi* could be the result of four mechanisms: automixis, sexuality (amphimixis), gene conversion, and mitotic recombination. It is likely that a combination of these not-mutually exclusive mechanisms has generated the convergence of nuclear gene copies, given the evidence of genetic exchange, even though asexual reproduction is the preponderant mode of reproduction in *T. cruzi*.

Divergence Times. Rate constancy is not rejected when a subset of *T. cruzi* and bat trypanosome TR or COII-ND1 sequences are analyzed with the sequences from *T. brucei* [COII-ND1: 2 Δ–ln (L) = 6.06, $P = 0.87$, 11 d.f; TR: 2 Δ–ln (L) = 6.78, $P = 0.82$, 11 d.f] or with *T. brucei* and *Leishmania* sequences [COII-ND1: 2 Δ–ln (L) = 6.65, $P = 0.92$, 13 d.f; TR: 2 Δ–ln (L) = 10.88, $P = 0.54$, 12 d.f]. The estimates of divergence time suggest that the extant lineages of *T. cruzi* originated during the Miocene or early Pliocene (Table 5), more recently than previously proposed (37–88 Myr) (43, 44) but consistent with the age of its triatomine vectors (45, 46). However, the COII-ND1 and TR estimates are notably different.

Discussion

The results provide evidence of hybridization between strains from two divergent groups of *T. cruzi*, demonstrate mitochondrial introgression across distantly related lineages, and reveal genetic exchange among closely related strains. Genetic exchange among closely related strains is also evidenced by additional data (unpublished results). However, genetic exchange among distantly related strains is either infrequent or a recent phenomenon, as otherwise the large number of fixed genetic differences among sequences from the major clades would not exist. Further, the observation of stable hybrid genotypes is consistent with previous conclusions about the preponderance of

asexual reproduction in *T. cruzi*. Our data do not reject the hypothesis of clonal population structure in *T. cruzi* (2, 3, 47) but suggest that genetic exchange has also played a significant role in generating genetic diversity in this parasite.

The genealogies presented agree with a study suggesting three major phylogenetic lineages in *T. cruzi* on the basis of sequence data from an intergenic nuclear region (48) and two unpublished studies using sequences from the mitochondrial *cytochrome b* gene and the nuclear rRNA promoter region (S. Brisse, G. Buck, M. de Carvalho, and M. Serrano, personal communication), and from the *18S*, *Tc52*, and *Gpi* genes (F. Tarrieu, H. Broutin, C. Barnabé, and B. Oury, personal communication). However, only our study provides clear evidence of the hybrid nature of isoenzyme types 43 and 39. The homozygosity observed in the other nuclear genes from isoenzyme types 43 and 39 is possibly the result of gene conversion, which is expected for the ribosomal genes.

The gene genealogies are not fully compatible with MLEE and RAPD phenograms (5, 6, 11). All strains from *T. cruzi* I (Z1) carry only alleles from sequence clade A, but the gene genealogies suggest that *T. cruzi* II (Z2, Z3) is paraphyletic because strains from this group carry alleles from clades B, C, and D, even though B is more closely related to A (Z1) than to C or D. The presence of the hybrid strains within *T. cruzi* II could

Table 5. Estimates of the age (Myr) of the most recent common ancestor of various trypanosomatids, with *T. brucei* or *Leishmania* as outgroups

	T. brucei	*Leishmania* *
COII-ND1		
T. cruzi clade	10.45 ± 2.27	16.54 ± 3.22
T. cruzi—bat trypanosomes	16.55 ± 3.50	25.06 ± 4.43
T. cruzi—*T. brucei*	100	100
T. cruzi—*Leishmania*	—	140.43 ± 35.25
TR		
T. cruzi clade	3.91 ± 1.35	3.20 ± 1.13
T. cruzi—bat trypanosomes	8.15 ± 1.97	6.88 ± 1.69
T. cruzi—*T. brucei*	100	100
T. cruzi—*Leishmania*	—	250.83 ± 71.45

The age of the *T. cruzi*–*T. brucei* clade is fixed at 100 Myr (34, 35).
Leishmania tarentolae COII-ND1 from GenBank (accession no. M10126); *L. braziliensis* COII-ND1 from this study (accession no. AF359057); *Leishmania donovani* TR from GenBank (accession no. Z23135).

Machado and Ayala

partially explain this inconsistency, because including them in the calculation of genetic distances may generate spurious relationships by increasing the homoplasy in the data, leading to the potential artifact of including isoenzyme types 30 and 32 within *T. cruzi* II instead of placing them as the most divergent group of *T. cruzi* strains. However, although homoplasy levels decrease after removing all MLEE data from isoenzyme types 39 and 43 (the consistency index changes from 0.592 to 0.727), cluster analyses of the reduced MLEE data do not recover the three clades revealed by the sequence data (results not shown).

An alternative possibility is that the recent ancestor of *T. cruzi* consisted of three or four isolated lineages (corresponding to the distinct sequence clades), but recent genetic exchanges resulted in some *T. cruzi* II strains that carry alleles from two or three ancestral lineages. This explanation assumes that despite the problems of higher homoplasy and uncertain homology of the MLEE and RAPD data, these data better represent the history of *T. cruzi* because they reflect genome-wide relationships and not just the relationships between alleles from one locus represented by the gene genealogies. If this explanation is correct, future sequencing studies should find more complex patterns of allele combinations.

Our results show that the strain CL Brener (43-CL F11F5), chosen for the *T. cruzi* genome project (see also refs. 11 and 42), is hybrid. This choice is fortunate in that it could yield sequence information from two of the four major clades of *T. cruzi*. But a greater sequencing effort will be required to complete the project, because each chromosome may need to be sequenced twice, given the large number of expected nucleotide differences between two haploid genomes, one from clade B and the other from clade C. The haploid genome size of strain CL Brener is 43 Mb (42). Assuming that the sequence divergence between homologous genes from clades B and C is 1.5% (Table 2), the two genomes will have 6.5×10^5 nucleotide differences on average. That might be an underestimate, because DHFR-TS and TR are conserved genes, but it could also be an overestimate, because gene conversion has probably homogenized the sequence of several genomic regions. Special attention should be given to determining the linkage group to which each sequenced clone corresponds. This may not be trivial given that divergence will greatly vary across different regions of the genome (especially if gene conversion has occurred), and because there is evidence of recombination in the hybrids (unpublished results),

which suggests that each chromosome from CL Brener is likely to be a mosaic of the two parental genotypes.

The date used for calibrating substitution rates in TR and COII-ND1 is a conservative estimate of the age of the common ancestor of mammalian trypanosomes (34, 35). It is consistent with biogeographical and phylogenetic evidence of a Gondwanan origin of trypanosomes (35) and a separation of the African and American trypanosomes (*T. brucei* and *T. cruzi*) at the time of separation of Africa and South America about 100 Myr ago (34). This calibration date postdates by about 40 Myr the first appearance in the fossil record of the Protoglossinae (49), the presumed ancestor of the tsetse fly, the current vector of *T. brucei*.

The estimated dates for the origin of *T. cruzi* seem reasonable. The estimated rate of substitution for the COII gene within the clade comprised by the bat trypanosomes and *T. cruzi* (2.12×10^{-8} substitutions per silent site per year) is slightly lower than the average silent substitution rates estimated for the COII gene of *Drosophila* (2.34×10^{-8}) or primates (3.51×10^{-8} and 2.54×10^{-8}, whether or not humans are included). Silent substitution rates estimated for TR and DHFR-TS by using the age of the common ancestor of bat trypanosomes and *T. cruzi* estimated with the COII-ND1 data (5.12×10^{-9} and 7.34×10^{-9}, respectively) or with the TR data (11.42×10^{-9} and 15.45×10^{-9}, respectively) fall within the range of silent rates estimated for mammalian and *Drosophila* nuclear genes (3.5×10^{-9} and 15.6×10^{-9} per silent site per year, respectively) (50). Further, a recent diversification of the major lineages of *T. cruzi* is consistent with the low level of divergence among *T. cruzi* sequences (Table 2) and is compatible with evidence that the triatomine vectors of the parasite (Hemiptera, Reduviidae, Triatominae) have recently evolved from predatory reduviid ancestors (45, 46). Older dates proposed for the origin of *T. cruzi* (37–88 Myr) (43, 44) are improbable. These dates lead to unlikely estimates of the origin of mammalian trypanosomes (at least 360 Myr) and to rates of silent substitution in *T. cruzi* that are at least one order of magnitude lower than those estimated for other organisms.

1. Miles, M. A., Souza, A., Povoa, M., Shaw, J. J., Lainson, R. & Toye, P. J. (1978) *Nature (London)* **272**, 819–821.
2. Tibayrenc, M., Ward, P., Moya, A. & Ayala, F. J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 115–119.
3. Tibayrenc, M. & Ayala, F. J. (1988) *Evolution (Lawrence, KS)* **42**, 277–292.
4. Souto, R. P. & Zingales, B. (1993) *Mol. Biochem. Parasitol.* **62**, 45–52.
5. Tibayrenc, M., Neubauer, K., Barnabé, C., Guerrini, F., Skarecky, D. & Ayala, F. J. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 1335–1339.
6. Souto, R. P., Fernandes, O., Macedo, A. M., Campbell, D. A. & Zingales, B. (1996) *Mol. Biochem. Parasitol.* **83**, 141–152.
7. Tibayrenc, M. (1995) *Adv. Parasitol.* **36**, 47–115.
8. Anonymous (1999) *Mem. Inst. Oswaldo Cruz* **94**, 429–432.
9. Brisse, S., Barnabé, C. & Tibayrenc, M. (2000) *Int. J. Parasitol.* **30**, 35–44.
10. Brisse, S., Dujardin, J. & Tibayrenc, M. (2000) *Mol. Biochem. Parasitol.* **111**, 95–105.
11. Brisse, S., Barnabé, C., Banuls, A. L., Sidibe, I., Noel, S. & Tibayrenc, M. (1998) *Mol. Biochem. Parasitol.* **92**, 253–263.
12. Oliveira, R. P., Broude, N. E., Macedo, A. M., Cantor, C. R., Smith, C. L. & Pena, S. D. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 3776–3780.
13. Povoa, M. M., de Souza, A. A., Naiff, R. D., Arias, J. R., Naiff, M. F., Biancardi, C. B. & Miles, M. A. (1984) *Ann. Trop. Med. Parasitol.* **78**, 479–487.
14. Carrasco, H. J., Frame, I. A., Valente, S. A. & Miles, M. A. (1996) *Am. J. Trop. Med. Hyg.* **54**, 418–424.
15. Bogliolo, A. R., Lauria-Pires, L. & Gibson, W. C. (1996) *Acta Trop.* **61**, 31–40.
16. Higo, H., Yanagi, T., Matta, V., Agatsuma, T., Cruz-Reyes, A., Uyema, N., Monroy, C., Kanbara, H. & Tada, I. (2000) *Parasitology* **121**, 403–408.
17. Woese, C. R., Gibson, J. & Fox, G. E. (1980) *Nature (London)* **283**, 212–214.
18. Dykhuizen, D. E. & Green, L. (1991) *J. Bacteriol.* **173**, 7257–7268.
19. Dykhuizen, D. E., Polin, D. S., Dunn, J. J., Wilske, B., Preac-Mursic, V., Dattwyler, R. J. & Luft, B. J. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 10163–10167.
20. Koufopanou, V., Burt, A. & Taylor, J. W. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 5478–5482.
21. White, M. J. D. (1945) *Animal Cytology and Evolution* (Cambridge Univ. Press, Cambridge, U.K.).
22. Suomalainen, E. (1950) *Adv. Genet.* **3**, 193–252.
23. Lokki, J. (1976) *Hereditas* **83**, 57–64.
24. Birky, C. W., Jr.. (1996) *Genetics* **144**, 427–437.
25. Welch, D. M. & Meselson, M. (2000) *Science* **288**, 1211–1215.
26. Sullivan, F. X. & Walsh, C. T. (1991) *Mol. Biochem. Parasitol.* **44**, 145–147.
27. Reche, P., Arrebola, R., Olmo, A., Santi, D. V., Gonzalez-Pacanowska, D. & Ruiz-Perez, L. M. (1994) *Mol. Biochem. Parasitol.* **65**, 247–258.
28. Kim, K. S., Teixeira, S. M., Kirchhoff, L. V. & Donelson, J. E. (1994) *J. Biol. Chem.* **269**, 1206–1211.
29. Swofford, D. L. (1998) PAUP*, *Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer, Sunderland, MA).
30. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
31. Tamura, K. & Nei, M. (1993) *Mol. Biol. Evol.* **10**, 512–526.
32. Goldman, N. (1993) *J. Mol. Evol.* **36**, 182–198.
33. Shimodaira, H. & Hasegawa, M. (1999) *Mol. Biol. Evol.* **16**, 1114–1116.
34. Lake, J. A., de la Cruz, V. F., Ferreira, P. C., Morel, C. & Simpson, L. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4779–4783.
35. Stevens, J. R., Noyes, H. A., Dover, G. A. & Gibson, W. C. (1999) *Parasitology* **118**, 107–116.
36. Felsenstein, J. (1988) *Annu. Rev. Genet.* **22**, 521–565.
37. Yang, Z. (1997) *Comput. Appl. Biosci.* **13**, 555–556.
38. Miyashita, N. T., Aguade, M. & Langley, C. H. (1993) *Genet. Res.* **62**, 101–109.
39. Yang, Z. (1994) *J. Mol. Evol.* **39**, 105–111.
40. Hasegawa, M., Kishino, H. & Yano, T. (1985) *J. Mol. Evol.* **22**, 160–174.
41. Hey, J. & Wakeley, J. (1997) *Genetics* **145**, 833–846.
42. Cano, M. I., Gruber, A., Vazquez, M., Cortes, A., Levin, M. J., Gonzalez, A., Degrave, W., Rondinelli, E., Zingales, B., Ramirez, J. L., et al. (1995) *Mol. Biochem. Parasitol.* **71**, 273–278.
43. Briones, M. R., Souto, R. P., Stolf, B. S. & Zingales, B. (1999) *Mol. Biochem. Parasitol.* **104**, 219–232.
44. Gaunt, M. & Miles, M. (2000) *Mem. Inst. Oswaldo Cruz* **95**, 557–565.
45. Gorla, D. E., Dujardin, J. P. & Schofield, C. J. (1997) *Acta Trop.* **63**, 127–140.
46. Schofield, C. (2000) *Mem. Inst. Oswaldo Cruz* **95**, 535–544.
47. Tibayrenc, M., Kjellberg, F. & Ayala, F. J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2414–2418.
48. Robello, C., Gamarro, F., Castanys, S. & Alvarez-Valin, F. (2000) *Gene* **246**, 331–338.
49. Lambrecht, F. L. (1980) *Proc. Am. Phil. Soc.* **124**, 367–385.
50. Li, W.-H. (1997) *Molecular Evolution* (Sinauer, Sunderland, MA).

EVOLUTION