

Genome-Scale Multilocus Microsatellite Typing of *Trypanosoma cruzi* Discrete Typing Unit I Reveals Phylogeographic Structure and Specific Genotypes Linked to Human Infection

Martin S. Llewellyn^{1*}, Michael A. Miles¹, Hernan J. Carrasco², Michael D. Lewis¹, Matthew Yeo¹, Jorge Vargas³, Faustino Torrico⁴, Patricio Diosque⁵, Vera Valente⁶, Sebastiao A. Valente⁶, Michael W. Gaunt¹

1 London School of Hygiene and Tropical Medicine, London, United Kingdom, **2** Instituto de Medicina Tropical, Universidad Central de Venezuela, Los Chaguaramos, Caracas, Venezuela, **3** Centro Nacional de Enfermedades Tropicales, Santa Cruz, Bolivia, **4** Universidad Mayor de San Simon, Cochabamba, Bolivia, **5** Consejo Nacional de Investigaciones Científicas y Técnicas, Instituto de Patología Experimental, Universidad Nacional de Salta, Salta, Argentina, **6** Instituto Evandro Chagas, Rodovia, Belem, Para, Brazil

Abstract

Trypanosoma cruzi is the most important parasitic infection in Latin America and is also genetically highly diverse, with at least six discrete typing units (DTUs) reported: Tc I, IIa, IIb, IIc, IId, and IIe. However, the current six-genotype classification is likely to be a poor reflection of the total genetic diversity present in this undeniably ancient parasite. To determine whether epidemiologically important information is “hidden” at the sub-DTU level, we developed a 48-marker panel of polymorphic microsatellite loci to investigate population structure among 135 samples from across the geographic distribution of TcI. This DTU is the major cause of resurgent human disease in northern South America but also occurs in silvatic triatomine vectors and mammalian reservoir hosts throughout the continent. Based on a total dataset of 12,329 alleles, we demonstrate that silvatic TcI populations are extraordinarily genetically diverse, show spatial structuring on a continental scale, and have undergone recent biogeographic expansion into the southern United States of America. Conversely, the majority of human strains sampled are restricted to two distinct groups characterised by a considerable reduction in genetic diversity with respect to isolates from silvatic sources. In Venezuela, most human isolates showed little identity with known local silvatic strains, despite frequent invasion of the domestic setting by infected adult vectors. Multilocus linkage indices indicate predominantly clonal parasite propagation among all populations. However, excess homozygosity among silvatic strains and raised heterozygosity among domestic populations suggest that some level of genetic recombination cannot be ruled out. The epidemiological significance of these findings is discussed.

Citation: Llewellyn MS, Miles MA, Carrasco HJ, Lewis MD, Yeo M, et al. (2009) Genome-Scale Multilocus Microsatellite Typing of *Trypanosoma cruzi* Discrete Typing Unit I Reveals Phylogeographic Structure and Specific Genotypes Linked to Human Infection. PLoS Pathog 5(5): e1000410. doi:10.1371/journal.ppat.1000410

Editor: Edward C. Holmes, The Pennsylvania State University, United States of America

Received: January 30, 2009; **Accepted:** April 1, 2009; **Published:** May 1, 2009

Copyright: © 2009 Llewellyn et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Funding was provided by a Wellcome Trust junior research fellowship (MWG), The European Union Seventh Framework Programme grant 223034 (MAM), The Dr. Gordon-Smith Scholarship (MSL), The Swire Charitable Trust (MSL), The De Laszlo Foundation (MSL), and FONACIT (Venezuela) project G-2005000827 (HJC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: martin.llewellyn@lshtm.ac.uk

Introduction

T. cruzi, the etiological agent of Chagas disease, is a vector borne zoonosis and considered the most important parasitic infection in Latin America. In excess of 10 million people are thought to carry the parasite, with ten times that number at risk (<http://www.who.int>). Consistent with a long history on the continent [1], *T. cruzi* ecology in the silvatic environment is highly complex. Over 73 mammalian genera and just over half of 137 described species of haematophagous triatomine bug are involved with parasite carriage and transmission [2,3]. *T. cruzi* has an endemic range that stretches from the Southern USA to Northern Argentina. Most human infection is found in Central and South America and occurs primarily through contact with the contaminated faeces of domiciliated triatomine vector species.

Genotypic data support the existence of six stable discrete typing units (DTUs) in *T. cruzi*: TcI, TcIIa, TcIIb, TcIIc, TcIId, and TcIIe

[4]. Greatest molecular divergence is observed between TcI and TcIIb [1,4]. TcIIa and TcIIc have distinct genotypes but their affinities to other DTUs are inadequately understood [4,5]. TcIId and TcIIe are hybrids, and have haplotypes shared across TcIIb and TcIIc [1,6]. The ecological and epidemiological relevance of different *T. cruzi* DTUs have been the subject of considerable debate. Using a retrospective analysis of all available genotype records, we recently showed that diversification in the silvatic environment is likely to be driven by ecological niche as well as host species, with arboreal Didelphimorpha (opossums) the principal hosts of TcI, and terrestrial Cingulata (armadillos) the principal hosts of TcIIc [7]. TcI is a major agent for human disease north of the Amazon Basin [8,9], but is also ubiquitous in silvatic transmission cycles throughout the Americas [10,11]. In the Southern Cone region of South America, DTUs TcIIb, TcIId, and TcIIe cause most human infection [10]. With the exception of putative epizootic outbreaks [12], TcIIb, TcIId, and TcIIe are so far rare in the silvatic cycle [7].

Author Summary

The arrival of the *Trypanosoma cruzi* online genome now provides vital information for the study of Chagas disease. Using this resource, we identified and developed a genome-scale panel of rapidly evolving microsatellite markers that can be used to unravel the micro-epidemiology of this parasite. We then tested these against a panel of isolates belonging to the most widely occurring and ancient major lineage, *T. cruzi* I (TcI). Our study includes samples from across the geographical distribution of this lineage, including isolates from wild vectors, domestic vectors, as well as wild mammalian reservoirs and human hosts. This is the first time *T. cruzi* has been subjected to such high-resolution population genetic analysis. Our study shows that important epidemiological information lies at the intra-lineage level, especially when wild and domestic populations of parasite are compared. Crucially, in Venezuela, where Chagas disease may be resurgent despite decades of control effort, genotypes of parasites found in the wild are rarely represented in humans, despite evidence that infected wild vectors do invade houses. In this manuscript, we examine the epidemiological implications of this finding and others, and suggest how the approach we have developed can now be used to investigate the true nature of parasite transmission at Chagas disease foci throughout the Americas.

The current six-genotype classification of *T. cruzi* is likely to provide a poor reflection of the total diversity present. Abundant evidence from nucleotide sequence [13,14], microsatellite [5,15], RAPD [16] and MLEE [11,17] data exists to suggest that considerable genetic variation is hidden at the sub-DTU level. Combining an adequate sample size with a genetic marker of sufficient resolution to unravel fine-scale relationships, however, remains a significant challenge. Indeed few, if any, detailed studies exist to document the population genetic diversity of a mammalian protozoan parasite in its true silvatic cycle. For many zoonotic infections, e.g. *Cryptosporidium* spp, *Trypanosoma brucei* spp, *Leishmania* spp, and *Toxoplasma gondii*, domestic mammals and (where applicable) associated vectors are the obvious target for population-level studies of parasite genetic variation since these are the most likely source of human outbreaks. For *T. cruzi*, this rationale must also extend to wild reservoir hosts. Many, especially opportunistic scavengers like *D. marsupialis*, also come into close contact with humans, either directly, or via infected silvatic vector species. In areas now free or without a history of vectorial domestic transmission, oral outbreaks are a growing concern [18].

High-resolution population genetic studies of other parasitic zoonoses have facilitated epidemiological tracking of human disease outbreaks, with obvious implications for the planning of effective disease control [19,20]. Molecular methods transformed our early understanding of *T. cruzi* epidemiology, with the revelation that distinct transmission cycles (domestic/silvatic) could harbour different major lineages of parasite [21]. Predominantly clonal propagation observed in *T. cruzi* is in keeping with this result, where micro-endemic clones with characteristic host propensities, geographic distribution, medical significance and biological attributes should exist within the parasite population [22]. However, widespread multi-host *T. cruzi* lineages like TcI persist outside of this paradigm. With the advent of the *T. cruzi* genome [23], the stage is now set to re-examine the micro-epidemiology of human disease outbreaks in TcI in the context of ultra-high resolution genetic analysis and, crucially, silvatic parasite populations. In this study we have developed a multilocus

microsatellite typing (MLMT) system for TcI and applied it to parasite isolates from throughout the Americas. While this is among the largest panel of isolates from a single DTU ever analysed, sample sizes are still restrictive. Similarly, widespread deviation from Mendelian sexuality in *T. cruzi* limits the inferences that can be made from standard population genetic analyses. To circumvent these issues, we largely avoided model-based population assignment protocols (e.g. Structure [24]). In spite of these limitations, we are able to identify key features of silvatic TcI populations and highlight population genetic processes that accompany a switch to the human host in two endemic areas. In doing so we show that the pattern of within-DTU parasite genetic diversity may contain vital epidemiological information in terms of control strategies, parasite pathogenesis and ultimately human disease.

Results

A final dataset comprising 12,329 alleles (excluding missing data) from 135 isolates was subjected to analysis. Most strains presented one or two alleles at each locus. Multiple (≥ 3) alleles were observed at a small proportion of loci (0.98%) and only among strains not biologically cloned. Multiclonality, rather than aneuploidy, was determined to be the major source of this phenomenon by reference to analysis of a subset of nine microsatellite loci across 211 clones taken from a subset of eight strains that demonstrated multiple alleles at individual loci in the uncloned state (data not shown). Samples were allocated to seven populations: North and Central American ($AM_{North/Cen}$), Venezuelan silvatic (VEN_{silv}), North Eastern Brazil ($BRAZ_{North-East}$), Northern Bolivia (BOL_{North}), Northern Argentina (ARG_{North}), Bolivian and Chilean Andes ($ANDES_{Bol/Chile}$) and Venezuelan domestic (VEN_{dom}). A full list of sample allocations is included in Table S2 and the rationale for the assignment of individuals to populations is detailed in the *Methods* section.

Genetic diversity and rare allele frequency distributions

Greatest genetic diversity was observed in populations drawn from palm and lowland moist forest associated ecotopes in VEN_{silv} , $BRAZ_{North-East}$ and BOL_{North} (Allelic richness (A_r) = 2.229–2.344, Table 1). Small, genetic-drift prone populations lose rare alleles at a faster rate than they can be replenished by mutation. Poisson-distributed rare allele frequency plots for VEN_{silv} , $BRAZ_{North-East}$ and BOL_{North} are, instead, characteristic of populations with a large, stable N_e at mutation-drift equilibrium (Figure S1) [25]. It is of note that patterns of both allelic richness and rare allele distribution are consistent across VEN_{silv} ($n = 37$) $BRAZ_{North-East}$ ($n = 39$) and BOL_{North} ($n = 16$), largely independent of sample size (Table 1, Figure S1). Additionally, the size of geographic focus had little relevance in determining the amount of diversity present in these populations. A marginal reduction in allelic richness, for example, was observed between $BRAZ_{North-East}$ and BOL_{North} ($A_r = 2.344$ – 2.229 , Table 1), despite a massive reduction in sampling area ($\sim 4,500,000$ km²– 10 km²).

A considerable reduction in diversity among silvatic isolates from $AM_{North/Cen}$ was observed with respect to VEN_{silv} , $BRAZ_{North-East}$ and BOL_{North} , again independent of sample size ($A_r = 1.532$, Table 1), concurrent with a reduction in rare allele frequency (Figure S1) and, assuming neutrality, implying that this population has been subject to a greater level of genetic drift in its recent past. Among three further populations, either exclusively comprised of domestic isolates (i.e. VEN_{dom}), or including a mixture of domestic and silvatic isolates (i.e. $ANDES_{Bol/Chil}$, ARG_{North}), a reduction in diversity was also observed ($A_r = 1.407$ – 1.794). Here, to varying

Table 1. Population genetic parameters for seven Tci populations.

Population	N/G	MNA	A _r ^a	H _O ^b	H _E ^b	% HD ^c	%HE ^d	I _A ^e	P-Value ^f
AM _{North/Cen}	7/7	1.92	1.532	0.332	0.445	0.00	0.00	2.39	0.005
VEN _{silv}	37/37	6.45	2.337	0.449	0.637	44.19	0.00	1.38	<0.001
BRAZ _{North-East}	39/39	6.67	2.344	0.383	0.571	50.00	0.00	2.03	<0.001
BOL _{North}	16/16	4.67	2.229	0.467	0.643	17.50	0.00	3.98	<0.001
ARG _{North}	10/10	2.41	1.794	0.535	0.551	8.82	2.94	12.37	<0.001
ANDES _{Bol/Chile}	11/11	1.73	1.407	0.406	0.396	3.85	7.69	2.05	<0.001
VEN _{dom}	13/13*	2.02	1.486	0.421	0.422	7.14	14.29	1.21	0.011

*Two samples included in Figure 1 were excluded from population analysis due to DNA availability issues and consequential high levels of missing data.

^aAllelic richness (sample size corrected).

^bMean observed and expected heterozygosity across all loci.

^cProportion of loci showing a significant deficit in heterozygosity after a sequential Bonferroni correction.

^dProportion of loci showing significant excess heterozygosity after a sequential Bonferroni correction.

^eThe Index of Association.

^fP-value for index of association, calculated by comparison to a null distribution of 1,000 randomised datasets.

N, number of isolates in population; G, number of multilocus genotypes per population; MNA, mean number of alleles per locus.

doi:10.1371/journal.ppat.1000410.t001

degrees, rare allele frequency plots again demonstrate a possible reduction in N_e by comparison to major silvatic populations (Figure S1).

Heterozygosity indices

High levels of genetic diversity in the principal silvatic populations sampled (VEN_{silv} , $BRAZ_{North-East}$ and BOL_{North}) gave rise to correspondingly large estimates of expected heterozygosity ($H_E = 0.571$ – 0.643 , Table 1). However, observed levels of heterozygosity were substantially lower than those expected under Hardy-Weinberg Equilibrium ($H_O = 0.383$ – 0.467 , Table 1) and statistical significance could be attached to this observation at the level of individual loci (Table 1). Silvatic isolates from $AM_{North/Cen}$ demonstrated similar heterozygous deficit over loci, but, owing to sample size constraints, the same effect was not statistically supported at individual loci. In contrast to exclusively silvatic populations, observed levels of relative heterozygosity ($H_O:H_E$, Table 1), were raised in populations that included domestic isolates, especially in VEN_{dom} ($0.421:0.422$) and $ANDES_{Bol/Chile}$ ($0.406:0.396$).

To ascertain whether within-population subdivision had any effect on estimates of heterozygosity (i.e. Wahlund effects [26]), a number of subpopulations were picked (Table S2), representing, as far as possible, ‘true’ populations in space and time and within which no statistically supported genetic subdivision was observed on the basis of individual pair-wise distance measures (<75% bootstrap support, Figure 1). If a Wahlund effect was in operation, hidden population subdivision would act to artificially decrease observed heterozygosity levels (increase F_{IS}). Mean F_{IS} estimates over loci across three silvatic populations, two from BOL_{North} and a further from VEN_{silv} , instead remained positive ($F_{IS} = 0.157$) with a 99% confidence interval (CI) of $0.042:0.288$ obtained by bootstrapping over loci, thus providing non-probabilistic support for the deficit of heterozygosity as observed previously among the populations from which they were drawn, but also suggesting limited evidence of a Wahlund effect. A similar analysis of VEN_{dom} and selected isolates from $ANDES_{Bol/Chile}$ returned a negative F_{IS} value ($F_{IS} = -0.157$), although with a larger 99% CI encompassing zero (CI = $-0.421:0.12$). A test for significant difference between F_{IS} values over loci between these sub-population groups ($BOL_{North} + VEN_{silv} > ANDES_{Bol} + VEN_{dom}$) generated by random shuffling of alleles between groups, was negative ($p = 0.0639$),

albeit marginally, but suggests that direct comparisons of overall heterozygosity levels between these population groups should be approached with caution.

F_{IS} values were also analysed by syntenous sequence fragment (SSF) (as defined by the CL-Brener genome project; no chromosomal assembly is currently available), of which nine are represented in our panel with ≥ 2 microsatellite loci (Table S3, Figure 2). Calculations included both large and small (‘true’) population groupings for comparison. Mean F_{IS} values per SSF were consistently positive across major silvatic populations BOL_{North} , VEN_{silv} & $BRAZ_{North-East}$. This provides support for heterozygote deficiency at the population level, but also for a consistent level of heterozygosity between fragments. The same is broadly true for $AM_{North/Cen}$, concomitant with an increase in error associated with both a reduction in genetic diversity and sample size. F_{IS} values for sub-population groupings from BOL_{North} (BOL_{North}^1 & BOL_{North}^2) and VEN_{silv} (VEN_{silv}^3) reflect those of their source populations. A marginal decrease in F_{IS} across some SSFs could be attributed to a Wahlund effect, and not uniquely to error, but major inconsistencies were not observed. In contrast, high inter-SSF variance was observed in both $ANDES_{Bol/Chil}$ and VEN_{dom} , and to a lesser extent ARG_{North} , with some strongly negative values regardless of an increase in error about the mean. These data provide support for a distinction between these populations and those exclusively from the silvatic environment. At the sub-population level, the exclusion of Chilean isolates from $ANDES_{Bol}$ did not have a major impact on the derived values, although error in this case was extremely high.

Pair-wise measure of genetic distance

Figure 1 shows a Neighbor-joining tree based on pair-wise D_{AS} measures between individual isolates. Good bootstrap support was found for the grouping of isolates from VEN_{dom} and $AM_{North/Cen}$ (88.5%), for subdivision within Argentinean isolates (100%), for subdivision within BOL_{North} (92.5%), as well as for the grouping of isolates obtained from the Bolivian and Chilean Andes. In the silvatic environment no clear diversification was observed by reservoir host, a phenomenon supported by a non-significant estimate of F_{ST} between *Didelphis sp.* and non-*Didelphis sp.* reservoir hosts in $BRAZ_{North-East}$ ($F_{ST} = 0.006$, $p = 0.594$). Sample size restricts similar comparisons in other silvatic populations.

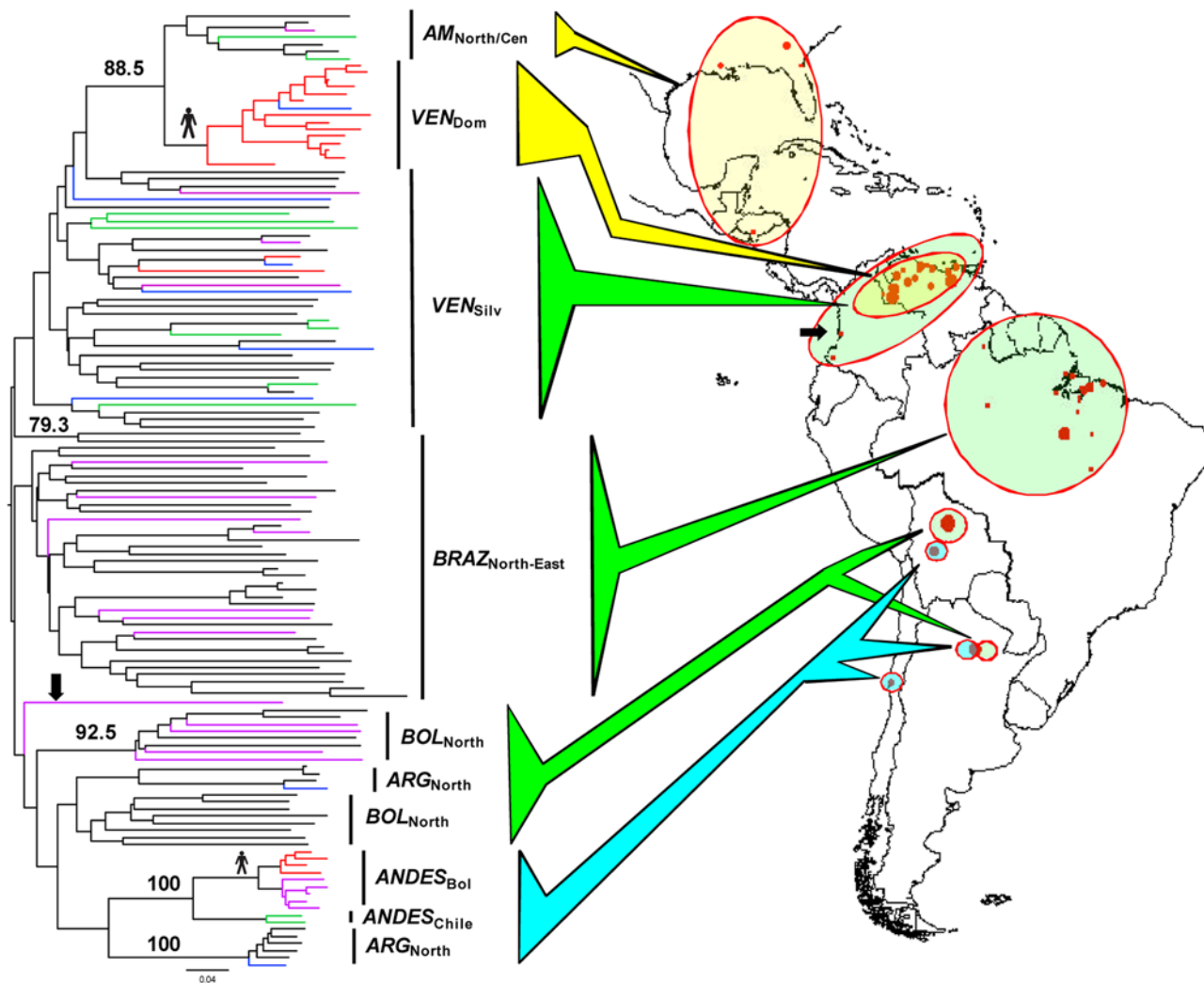


Figure 1. Unrooted neighbour-joining D_{AS} tree showing TcI population structure across the Americas. Based on the multilocus microsatellite profiles of 135 TcI isolates. D_{AS} values were calculated as the mean across 1,000 random diploid re-samplings of the dataset to accommodate multi-allelic loci. The presence of more than two alleles per locus did not disrupt the delineation of major clades (>90% majority consensus support). D_{AS} -based bootstrap values were calculated over 10,000 trees from 100 re-sampled datasets, and those >75% are shown on major clades. Branch colour codes indicate strain origin. Black: *Didelphis* species; purple: non-*Didelphis* mammalian reservoir; green: silvatic triatomine; red: human; blue: domestic triatomine. Colored block arrows and circles indicate broad population types. Yellow: Venezuelan domestic and North/Central American groups; green: major silvatic populations; blue: South-Western clade. Black arrow indicates Colombian outlier assigned to Brazilian population. Human symbol indicates putative genetic association with domestic transmission. Closed red circle area is proportionate to sampling density. See text for details of population codes.
doi:10.1371/journal.ppat.1000410.g001

A portion of the pair-wise genetic diversification observed in the dataset could be attributed to isolation by distance (IBD). A Mantel's test for matrix correspondence between pair-wise genetic (D_{AS}) and geographic distance (km) revealed a highly significant positive correlation between these two measures ($R_{XY} = 0.394$, $p < 0.0001$, Figure 3). Nonetheless, pair-wise comparisons also revealed considerable diversification between isolates from the same site in some instances (e.g. BOL_{North} - Mean $D_{AS} = 0.479 \pm 0.009$ (Standard Error)). Additionally, a number of outliers, representing comparisons within and between some groups of samples, are seen in Figure 3. These correspond to geographically disperse but relatively genetically homogeneous groups. Of particular interest are domestic isolates from Venezuela (VEN_{dom}), comparisons between which lie within the dashed box labelled 'D' in Figure 3. No significant IBD is observed among these isolates when analysed separately ($R_{XY} = 0.225$, $p = 0.0531$) in contrast to

those from the silvatic environment, which do show significant IBD (VEN_{silv} (Colombian outlier excluded, see Table S2) - $R_{XY} = 0.292$, $p = 0.0001$). A related observation is made among isolates from $AM_{North/Cen}$, where no significant IBD ($R_{XY} = 0.360$, $p = 0.161$) is observed. Again, these isolates fall as outliers in Figure 3 (Box B).

Population subdivision

Despite evidence of spatial structure across Amazonia at an individual level (Mantel's test VEN_{silv} , $BRAZ_{North-East}$, and BOL_{North} combined - $R_{XY} = 0.533$, $p < 0.0001$) the level of subdivision between these populations was generally low ($F_{ST} = 0.108$ – 0.148 , Table S1). Another observation not wholly consistent with IBD was a significant degree of subdivision between isolates from $ANDES_{Bol/Chil}$ and BOL_{North} ($F_{ST} = 0.304$) as compared with the strong connectivity between BOL_{North} and

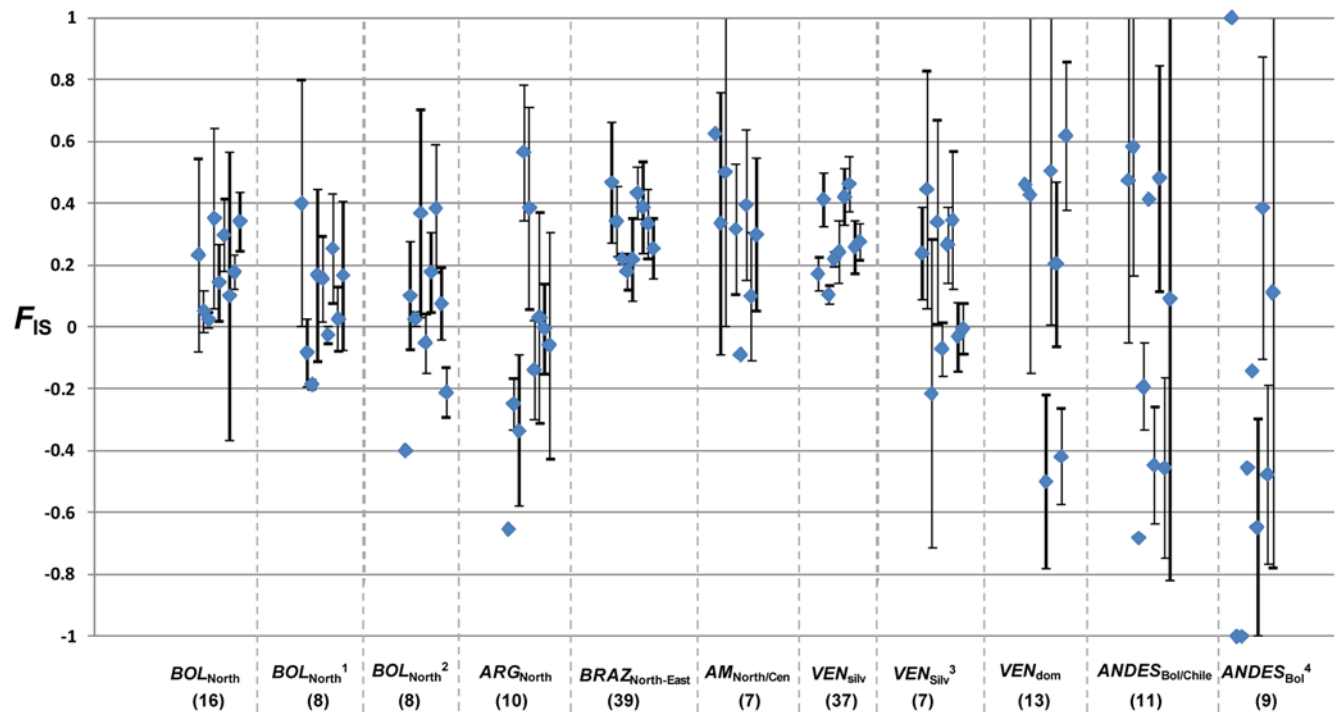


Figure 2. Mean F_{IS} values across loci on nine syntenous sequence fragments (SSFs) examined in eleven populations. Values suggest that gene conversion is a genomically diffuse process in homozygous silvatic populations. Error bars represent \pm standard error about the mean. Values without error bars correspond to SSFs containing only a single variable locus. Missing values correspond to SSFs containing no variable loci. Populations with postfix ^{1,2,3,4} are subsamples of larger populations. Numbers in parentheses indicate population size (n).
doi:10.1371/journal.ppat.1000410.g002

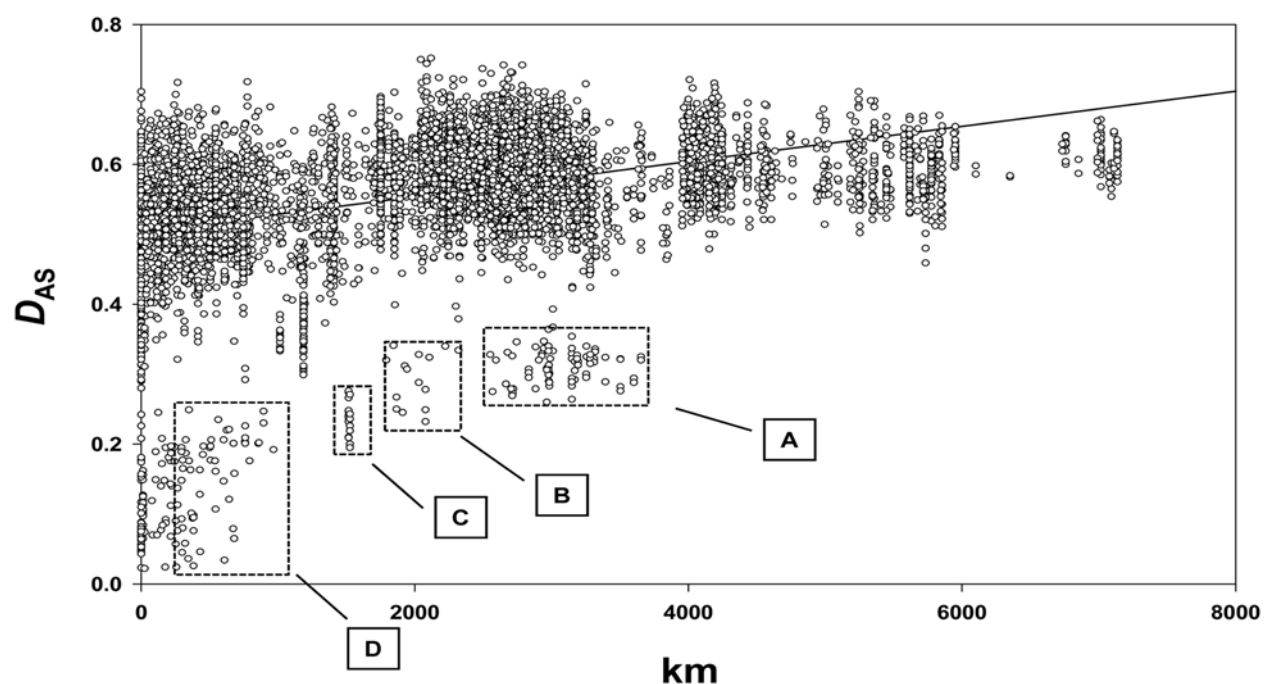


Figure 3. Continental scale spatial genetic structure among 135 TcI isolates from across the Americas. The graph shows a comparison between genetic (D_{AS}) and geographic (km) distance across the entire dataset. Each data point represents a comparison between two isolates, and there are thus 9,180 in total. A significant positive correlation between these two measures was observed ($R_{XY} = 0.394$, $p < 0.0001$). Outliers are highlighted by dashed lines. A – VEN_{dom} vs $AM_{North/Cen}$; B – $AM_{North/Cen}$ vs $AM_{North/Cen}$; C – $ANDES_{Bol}$ vs $ANDES_{Chile}$; D – VEN_{dom} vs VEN_{dom} . See text for details of population codes.
doi:10.1371/journal.ppat.1000410.g003

more distant lowland populations (e.g. BOL_{North} - VEN_{silv} F_{ST} = 0.148, Table S1). Most striking was the high level of discontinuity implied by the F_{ST} estimate between populations VEN_{dom} and VEN_{silv} (F_{ST} = 0.295), which approximately overlap in their distribution. To place this observation in context, similar subdivision is seen between populations VEN_{silv} and ARG_{North} (F_{ST} = 0.226) which lie >5000 km apart.

Linkage disequilibrium in TcI populations

Accounting for known physical linkage and excluding loci of unknown linkage group, the level of multilocus linkage disequilibrium was assessed using the I_A , and was found to be statistically greater than a null distribution generated from 1000 random permutations in all populations (Table 1). Thus, the current dataset is consistent with predominant clonality in this parasite.

Discussion

This study represents the most comprehensive attempt to document within-DTU diversity in *T. cruzi* to date. Nonetheless, some sample sizes remain limiting in population genetic terms, although efforts were made to correct for any confounding effects. Similarly, caution is required given the deviation of *T. cruzi* from the assumptions of most standard population genetic models due to clonality. Certainly, high levels of genetic diversity in the principal silvatic TcI populations examined in this study are consistent with the putative ancient (3–16 MYA) origin of this DTU [1]. Similarly, rare allele frequency plots are consistent with a large, stable N_e [25]. Furthermore, we have shown that similar diversity indices could be derived from a study area of 10 km² (BOL_{North}) as from one of 4,500,000 km² ($BRAZ_{North-East}$), which suggests that this study has barely scraped the surface of the total circulating diversity present. In the silvatic environment, no apparent component of this diversity is partitioned by host. Thus, a constrained, extant co-evolutionary relationship is not compatible with the current dataset; contrary to a recent study using mini-exon sequence data from a limited number of *Didelphis* TcI strains [13]. Previously, we have suggested the ecological niche, rather than reservoir host, plays the dominant role in driving *T. cruzi* diversification [7]. This reflects a current model for wider trypanosome evolution, where “ecological host-fitting” is thought to define parasite clades [27]. Low levels of subdivision (F_{ST}) between three populations sharing a similar ecotone across Amazonia are consistent with this supposition. While we demonstrate that TcI is eclectic in terms of host in arboreal lowland silvatic cycles, significant documentary evidence exists to suggest that *D. marsupialis* is the major carrier throughout much of lowland tropical South and Central America [7]. The majority of isolates examined here originate from this host. Tolerance by this species of high circulating parasitemia [28], as well as a possible propensity for non-vectorial transmission via infected territorial anal scent gland secretions [29], may predispose *D. marsupialis* to particularly intense *T. cruzi* transmission. Nonetheless, numerous vectors and secondary hosts are also implicated in TcI transmission and carriage [7,30], and parasite dispersal between geographic foci is unlikely to be linked to *D. marsupialis* alone. Continental scale spatial structure in silvatic TcI (Figure 3) fits with the general ecology of undisturbed wild transmission. Most triatomine vectors, for example, are ill-adapted to long-range flight, and are thus incapable of rapid parasite dispersal between distant foci, providing ample time for spatial differentiation to occur among parasite populations.

Sample size corrected genetic diversity estimates suggest a considerable reduction in genetic differentiation in $AM_{North/Cen}$

with respect to core silvatic populations. Furthermore, IBD breaks down among these isolates and a loss of rare alleles in this population could be interpreted as evidence of a recent population bottleneck [25]. Until recently, genetic studies of TcI diversity have failed to detect the signature of a rapid biogeographic expansion of this DTU into the USA [31]. Our findings are bolstered by low genetic diversity identified among new mini-exon sequence data derived from North and Central American TcI isolates [13], but greater sampling from this region would confirm our observations. The expansion of TcI into North and Central America is likely to have occurred since the formation of the Isthmus of Panama 2–4 MYA, providing a useful phylogeographic calibration point for future studies, and may correspond to the northerly migration of didelphid marsupials [32].

In this study, TcI strains from infected humans were sampled widely in Venezuela (Table S2). Although their sample size is currently limited (n = 15 for the domestic clade – includes one vector isolate (Table S2)), their robust genetic clustering, by comparison to the extensively sampled and genetically diverse parasite population from the silvatic environment, serves to make them representative and important. There are suggestions that Chagas disease is locally resurgent [33], and genetic discontinuity between the domestic population and most silvatic isolates raises significant questions regarding human disease transmission. Molecular data from the low-lying west of the country demonstrates that most silvatic and domestic populations of the principal vector, *Rhodnius prolixus*, are indistinguishable [34] and it follows that the parasite should also be invasive. However, in our study, the predominant *T. cruzi* strains infecting humans in the same and nearby areas bear little resemblance to those in the silvatic environment. Intriguingly, however, silvatic TcI genotypes prevail among almost all adult intradomiciliary triatomines sampled. All three triatomine species, *Triatoma maculata*, *Panstrongylus geniculatus*, and *R. prolixus* are also described from the silvatic environment in Venezuela [3] and could, therefore, be invasive, and the parasite strains infecting them not of human origin.

The occurrence of a domestic TcI clade in Venezuela, in spite of the presence of silvatic strains inside houses, presents an interesting problem. Among African trypanosomes (*T. brucei* spp.), human infective forms display only a limited array of genotypes (*T. b. rhodesiense* & *gambiense* [20,35]). Detailed studies of *T. b. brucei* population genetics in the silvatic environment are, however, lacking. Some evidence suggests that vectors and domestic mammalian reservoirs in *T. b. brucei* populations sympatric with human *T. b. rhodesiense* outbreaks support a greater diversity of strains [20]. However, no specific genes associated with human infectivity are known in *T. cruzi*, unlike in *T. b. rhodesiense* [36], that might drive the domestic expansion of an epidemic clone. Furthermore, silvatic-type TcI strains were capable of sustaining long-term, symptomatic infection in a subset of patients studied (Table S2). One possible confounder in our sampling, as in a recent population study of strains from West African *T. b. gambiense* symptomatic human infections [35], is a lack of samples from asymptomatic patients, which are required to refute an association between parasite genotype and virulence or pathogenicity.

In the absence of a clear adaptive explanation for the lack of diversification among Venezuelan domestic isolates on the basis of current data, an ecological one may be more parsimonious. Low transmission of the parasite to the human host by invasive adult triatomines may reflect the inefficient stercorarian route by which *T. cruzi* is normally spread [2]. Instead, repeated blood meals taken by domestic triatomine colonies may be necessary to ensure infective contact with the human host. In this case, other humans or domestic reservoirs will be the primary sources of human

infection, human and domestic vector migration the main driver of parasite dispersal, and a widespread, uniform domestic parasite genotype the result. This is an observation supported by a lack of IBD among domestic strains. The distribution of this genotype may be wider than described here, and there is now preliminary mini-exon sequence evidence that a domestic TcI genotype may also occur in Colombia [14].

The origin of the divergent Venezuelan human TcI population remains enigmatic. Isolates bear closest resemblance, by all measures employed in this study, to the North and Central American clade. In all likelihood, TcI populations migrated to the North prehistorically in conjunction with invasive mammalian reservoir hosts during the Great American Interchange [32]. Low genetic diversity is also identified in domestic *R. prolixus* populations from Central America [37], although presumably their northerly migration occurred many thousands of years later alongside human populations. It is highly improbable that domestic TcI strains carried northwards with *R. prolixus* subsequently dispersed so widely into the silvatic environment. The source of the domestic outbreak identified here probably remains sequestered among silvatic transmission cycles somewhere in the northerly distribution of TcI in South America.

A greater sampling effort is required around Cochabamba (*ANDES_{Bol}*) from both human and wild reservoirs before satisfactory conclusions can be drawn regarding local parasite transmission. Intriguingly, temporal heterogeneity seems to be negligible, and ~20 years separate the isolation of human and rodent strains (Table S2). Epidemiologically, congruence between populations from these two hosts is not unexpected. Local domestic and silvatic *T. infestans* populations match genetically and morphologically [38], and rodent isolates were collected within two kilometres of a major suburb of Cochabamba, where active urban transmission still occurs [39]. It is not clear, however, whether the parasite is invasive to the domestic setting, or whether domestic strains have re-invaded the silvatic cycle.

A major observation of this study, and in others examining genetic diversity in *T. cruzi* [1,4,15], is the deficiency of heterozygosity with respect to Hardy-Weinberg expectations observed in most populations. Similar observations are frequently made in the *Leishmania spp.* populations [40–42]. These levels of homozygosity are atypical with respect to other clonally reproducing diploids [35,43,44], where diversity is known to accumulate between alleles within the individual in the absence of recombination, leading to extreme levels of heterozygosity at homologous loci (the ‘Meselson effect’ [45]). Heterozygous deficiency in silvatic populations in our dataset cannot be uniquely attributed to hidden subdivision (Wahlund effect). We still find positive F_{IS} values in non-subdivided sub-samples of isolates within populations. Here, some increase in heterozygosity was observed (Figure 2), but not to the extent predicted by the Meselson effect. Multilocus linkage disequilibrium suggests that recombination is at most infrequent in the current dataset, although the Index of Association [46] is a relatively insensitive measure [44]. Thus, widespread loss of heterozygosity due to homologous recombination or gene conversion, not inbreeding, is the most likely genetic phenomenon that would result in the observed diversity in our data. Importantly, we can show that these events are apparently genomically diffuse, in silvatic populations at least. Most SSFs show similar levels of heterozygosity within populations, rather than some showing strong evidence of the Meselson effect (strongly negative F_{IS}) and others showing complete homozygosity, as would be expected of larger scale effects like ploidy cycles [47] or those following genome fusion events in yeast [48].

Populations *ANDES_{Bol/Chil}* and *VEN_{Dom}* share many features in population genetic terms: reduced diversity; non-equilibrium rare

allele frequencies; and high inter-SSF variance in F_{IS} values where strongly negative values on some SSFs reflect marginally raised overall heterozygosity at the population level. It remains to be seen whether these are unique characteristics of human TcI clades, whether they reflect possible past recombination events or some form of balancing selection, and we could not attribute significance to a decrease in F_{IS} from background levels. DTUs TcIId and TcIIe both show fixed heterozygosity at most loci because they are almost certainly hybrids [1,5], not due to the Meselson effect, and far in excess of heterozygosity levels observed in our dataset. Confirmation of the characteristics we have observed will come with more intensive sampling from domestic foci in both regions, as well as others across South America. Our data now show, with increasing support from other studies [13,14,49,50], that most *T. cruzi* lineages actually represent highly heterogeneous populations across their distribution, heterogeneity that may be highly informative in epidemiological terms. Control strategies would now greatly benefit from high density parasitological surveys in and around individual endemic disease foci, especially if a pathogenic human TcI genotype does exist, signalling a return in study design, if not methodology, to the early investigations of the 1970s [21]. Such studies should include parasite samples from silvatic mammals and vectors, as well as domestic sources, including both symptomatic and asymptomatic (or indeterminate) human cases. To this extent, using microsatellite markers developed here, *T. cruzi* population genetics can be observed at the finest scale and provide real insights into the true nature of Chagas disease transmission.

Methods

We assembled a panel of 135 *T. cruzi* samples belonging to TcI from throughout the silvatic distribution of this lineage (Table S2). DTU-level genotyping was achieved through analysis of the non-transcribed spacer region of the mini-exon gene, as described previously [51]. Microsatellite motifs were extracted from the draft sequence of the *T. cruzi* genome available at <http://www.genedb.org>. Four Mb of sequence, including at least 13 syntenous sequence fragments, were scanned for di- and tri-nucleotide repeats using a pattern matching script written in *sed*. An extension of the algorithm was included to extract the up and downstream flanking regions of the microsatellite sequence (~200 bp). Primer design was achieved in PRIMER3 [52].

Among 200 microsatellite loci identified, 45 were polymorphic. A further three were included from two previous studies [6,15]. Primers and binding sites are listed in Table S3. The following reaction cycle was implemented across all loci: a denaturation step of 4 minutes at 95°C, then 30 amplification cycles (95°C for 20 seconds, 57°C for 20 seconds, 72°C for 20 seconds) and a final 20 minute elongation step at 72°C. With a final volume of 10 µl, 1× ThermoPol Reaction Buffer (New England Biolabs (NEB), UK), 4 mM MgCl₂, 34 µM dNTPs; 0.75 pmols of each primer, 1 unit of *Taq* polymerase (NEB, UK) and 1 ng of genomic DNA were added. Five fluorescent dyes were used to label forward primers – 6-FAM & TET (Prologo, Germany), NED, PET & VIC (Applied Biosystems, UK). Allele sizes were determined using an automated capillary sequencer (AB3730, Applied Biosystems, UK), manually checked for errors and typed “blind” to control for user bias.

Microsatellite diversity analysis

Allelic richness estimates were calculated in FSTAT 2.9.3.2 [53] and corrected for sample size using Hurlbert’s rarefaction method [54] in MolKin v3.0 [55]. Pair-wise estimates of population

subdivision (F_{ST} , Table S1) and heterozygosity indices (Table 1) were estimated in ARLEQUIN 3.0 [56]. P-values for multiple tests were corrected using a sequential Bonferroni correction [57]. F_{IS} provides an alternative measure of heterozygosity by assessing the level of identity of alleles within individuals compared to that between individuals where +1 represents all individuals homozygous for different alleles, and -1 all individuals heterozygous for the same alleles. Mean F_{IS} estimates over loci in selected groups of sub-populations were calculated in FSTAT 2.9.3.2 using Weir and Cockerman's (1984) unbiased estimators [58]. Confidence intervals for F_{IS} estimates were calculated by bootstrapping over loci and tests for significant differences between values also in FSTAT 2.9.3.2 using 10,000 random permutations. Mean F_{IS} values per sequence fragment per population were calculated across standard (not Weir and Cockerman's) F_{IS} values in FSTAT 2.9.3.2. To assess the level of multilocus linkage disequilibrium, the Index of Association (I_A , multilocus) was calculated in MULTILOCUS 1.3b [46,59] (Table 1). Genetic distances between isolates were evaluated in MICROSAT under an infinite alleles model of microsatellite evolution using D_{AS} (1-proportion of shared alleles at all loci / n) [60] (Figure 1). To accommodate multi-allelic loci, a script was written in Microsoft Visual Basic to make multiple random diploid re-samplings of each multilocus profile (software available on request). Individual-level genetic distances were calculated as the mean across multiple re-sampled datasets. A single randomly sampled dataset was used for population-level analysis. A Mantel's test for matrix correspondence was executed in GENALEX 6 to compare pair-wise geographical (km) and genetic distance (D_{AS}) [61] (Figure 3). Samples were assigned to populations on an *a priori* basis according to geography and transmission cycle. D_{AS} - defined sample clustering was also used to inform population identity, and obvious outliers assigned to the correct genetic group (Figure 1). Rare allele frequency plots were

calculated as in Luikart *et al.*, 1998 [25], to detect perturbation following putative population events (e.g. population bottlenecks).

Supporting Information

Figure S1 Allele frequency classes among seven TcI populations.

Found at: doi:10.1371/journal.ppat.1000410.s001 (6.79 MB TIF)

Table S1 F_{ST} estimates of interpopulation differentiation for seven TcI subpopulations based on microsatellite data.

Found at: doi:10.1371/journal.ppat.1000410.s002 (0.04 MB DOC)

Table S2 Panel of *T. cruzi* TcI genotype isolates assembled for microsatellite analysis.

Found at: doi:10.1371/journal.ppat.1000410.s003 (0.31 MB DOC)

Table S3 Microsatellite loci and primers employed in this study.

Found at: doi:10.1371/journal.ppat.1000410.s004 (0.12 MB DOC)

Acknowledgments

Additional samples were provided by C. Barnabe at the IRD, Montpellier, France. Support in the field was provided by the following: Bolivia: M. R. Cortez, M. Solano, and B. Chang; Venezuela: D. Feliciangeli and M. Segovia. J. Rivett-Carnac designed the diploid re-sampling software.

Author Contributions

Conceived and designed the experiments: MSL MAM MWG. Performed the experiments: MSL HJC. Analyzed the data: MSL MDL. Contributed reagents/materials/analysis tools: HJC MDL MY JV FT PD VCV SAV. Wrote the paper: MSL MAM MDL MY.

References

- Machado CA, Ayala FJ (2001) Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. *Proc Natl Acad Sci U S A* 98: 7396–7401.
- Hoare CA (1972) The trypanosomes of mammals. Blackwell Scientific Publications. 749 p.
- Lent H, Wygodzinsky P (1979) Revision of the Triatominae and their significance as vectors of Chagas disease. *Bull Am Mus Nat Hist* 163: 123–520.
- Westenberger SJ, Barnabe C, Campbell DA, Sturm NR (2005) Two Hybridization Events Define the Population Structure of *Trypanosoma cruzi*. *Genetics* 171: 527–543.
- de Freitas JM, Augusto-Pinto L, Pimenta JR, Bastos-Rodrigues L, Goncalves VF, et al. (2006) Ancestral Genomes, Sex, and the Population Structure of *Trypanosoma cruzi*. *PLoS Pathogens* 2: e24. doi:10.1371/journal.ppat.0020024.
- Gaunt MW, Yeo M, Frame IA, Stothard JR, Carrasco HJ, et al. (2003) Mechanism of genetic exchange in American trypanosomes. *Nature* 421: 936–939.
- Yeo M, Acosta N, Llewellyn M, Sanchez H, Adamson S, et al. (2005) Origins of Chagas disease: Didelphis species are natural hosts of *Trypanosoma cruzi* I and armadillos hosts of *Trypanosoma cruzi* II, including hybrids. *Int J Parasitol* 35: 225–233.
- Miles MA, Cedillos RA, Pova MM, de Souza AA, Prata A, et al. (1981) Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas' disease? *Lancet* 1(8234): 1338–1340.
- Anez N, Crisante G, da Silva FM, Rojas A, Carrasco H, et al. (2004) Predominance of lineage I among *Trypanosoma cruzi* isolates from Venezuelan patients with different clinical profiles of acute Chagas' disease. *Trop Med Int Health* 9: 1319–1326.
- Miles MA, Yeo M, Gaunt M (2003) Genetic diversity of *Trypanosoma cruzi* and the epidemiology of Chagas disease. In: Kelly JM, ed. *Molecular mechanisms in the pathogenesis of Trypanosoma cruzi*. New York: Kluwer Academic/Plenum.
- Barnabe C, Brisse S, Tibayrenc M (2000) Population structure and genetic typing of *Trypanosoma cruzi*, the agent of Chagas disease: A multilocus enzyme electrophoresis approach. *Parasitology* 120: 513–526.
- Lisboa CV, Mangia RH, Rubiao E, de Lima NR, das Chagas Xavier SC, et al. (2004) *Trypanosoma cruzi* transmission in a captive primate unit, Rio de Janeiro, Brazil. *Acta Trop* 90: 97–106.
- O'Connor O, Bosseno MF, Barnabe C, Douzery EJ, Breniere SF (2007) Genetic clustering of *Trypanosoma cruzi* I lineage evidenced by intergenic minicodon gene sequencing. *Infect Genet Evol* 7: 587–593.
- Herrera C, Bargues MD, Fajardo A, Montilla M, Triana O, et al. (2007) Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia. *Infect Genet Evol* 7: 535–539.
- Oliveira RP, Broude NE, Macedo AM, Cantor CR, Smith CL, et al. (1998) Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. *Proc Natl Acad Sci U S A* 95: 3776–3780.
- Carrasco HJ, Frame IA, Valente SA, Miles MA (1996) Genetic exchange as a possible source of genomic diversity in sylvatic populations of *Trypanosoma cruzi*. *Am J Trop Med Hyg* 54: 418–424.
- Higo H, Miura S, Horio M, Mimori T, Hamano S, et al. (2004) Genotypic variation among lineages of *Trypanosoma cruzi* and its geographic aspects. *Parasitol Int* 53: 337–344.
- Coura JR, Junqueira AC, Fernandes O, Valente SA, Miles MA (2002) Emerging Chagas disease in Amazonian Brazil. *Trends Parasitol* 18: 171–176.
- Morrison LJ, Mallon ME, Smith HV, MacLeod A, Xiao L, et al. (2008) The population structure of the *Cryptosporidium parvum* population in Scotland: A complex picture. *Infect Genet Evol* 8: 121–129.
- MacLeod A, Tweedie A, Welburn SC, Maudlin I, Turner CM, et al. (2000) Minisatellite marker analysis of *Trypanosoma brucei*: Reconciliation of clonal, panmictic, and epidemic population genetic structures. *Proc Natl Acad Sci U S A* 97: 13442–13447.
- Miles M, Toye P, Oswald S, Godfrey D (1977) The identification by isoenzyme patterns of two distinct strain-groups of *Trypanosoma cruzi*, circulating independently in a rural area of Brazil. *Trans R Soc Trop Med Hyg* 71: 217–225.
- Tibayrenc M, Ayala FJ (2002) The clonal theory of parasitic protozoa: 12 years on. *Trends Parasitol* 18: 405–410.
- El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, et al. (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309: 409–415.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *J Hered* 89: 238–247.

26. Wahlund S (1928) Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* 11: 65–106.
27. Hamilton PB, Gibson WC, Stevens JR (2007) Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. *Mol Phylogenet Evol* 44: 15–25.
28. Legey AP, Pinho AP, Xavier SC, Marchevsky R, Carreira JC, et al. (2003) *Trypanosoma cruzi* in marsupial didelphids (*Philander fenestratus* and *Didelphis marsupialis*): Differences in the humoral immune response in natural and experimental infections. *Rev Soc Bras Med Trop* 36: 241–248.
29. Carreira JC, Jansen AM, de Nazareth Meirelles M, Costa e Silva F, Lenzi HL (2001) *Trypanosoma cruzi* in the scent glands of *Didelphis marsupialis*: the kinetics of colonization. *Exp Parasitol* 97: 129–140.
30. Gaunt M, Miles M (2000) The ecotopes and evolution of triatomine bugs (triatominae) and their associated trypanosomes. *Mem Inst Oswaldo Cruz* 95: 557–565.
31. Barnabe C, Yaeger R, Pung O, Tibayrenc M (2001) *Trypanosoma cruzi*: A considerable phylogenetic divergence indicates that the agent of Chagas disease is indigenous to the native fauna of the United States. *Exp Parasitol* 99: 73–79.
32. Marshall L, Sempere T (1993) Evolution of the Neotropical Cenozoic land mammal fauna in its geochronologic, stratigraphic, and tectonic context. In: Goldblatt P, ed. *Biological relationships between Africa and South America*. New Haven (Connecticut): Yale University Press. pp 329–392.
33. Feliciangeli MD, Campbell-Lendrum D, Martinez C, Gonzalez D, Coleman P, et al. (2003) Chagas disease control in Venezuela: Lessons for the Andean region and beyond. *Trends Parasitol* 19: 44–49.
34. Fitzpatrick S, Feliciangeli MD, Sanchez-Martin MJ, Monteiro FA, Miles MA (2008) Molecular Genetics Reveal That Silvatic *Rhodnius prolixus* Do Colonise Rural Houses. *PLoS Negl Trop Dis* 2: e210. doi:10.1371/journal.pntd.0000210.
35. Koffi M, De Meeus T, Bucheton B, Solano P, Camara M, et al. (2009) Population genetics of *Trypanosoma brucei gambiense*, the agent of sleeping sickness in Western Africa. *Proc Natl Acad Sci U S A* 106: 209–214.
36. Gibson W (2002) Will the real *Trypanosoma brucei rhodesiense* please step forward? *Trends Parasitol* 18: 486–490.
37. Monteiro FA, Barrett TV, Fitzpatrick S, Cordon-Rosales C, Feliciangeli D, et al. (2003) Molecular phylogeography of the Amazonian Chagas disease vectors *Rhodnius prolixus* and *R. robustus*. *Mol Ecol* 12: 997–1006.
38. Cortez MR, Emperaire L, Piccinalli R, Gurtler RE, Torrico F, et al. (2007) Sylvatic *Triatoma infestans* (Reduviidae, Triatominae) in the Andean valleys of Bolivia. *Acta Trop* 102: 47–54.
39. Medrano-Mercado N, Ugarte-Fernandez R, Butron V, Uber-Busek S, Guerra HL, et al. (2008) Urban transmission of Chagas disease in Cochabamba, Bolivia. *Mem Inst Oswaldo Cruz* 103: 423–430.
40. Schwenkenbecher JM, Wirth T, Schnur LF, Jaffe CL, Schallig H, et al. (2006) Microsatellite analysis reveals genetic structure of *Leishmania tropica*. *Int J Parasitol* 36: 237–246.
41. Kuhls K, Chicharro C, Canavate C, Cortes S, Campino L, et al. (2008) Differentiation and Gene Flow among European Populations of *Leishmania infantum* MON-1. *PLoS Negl Trop Dis* 2: e261. doi:10.1371/journal.pntd.0000261.
42. Al-Jawabreh A, Diezmann S, Muller M, Wirth T, Schnur LF, et al. (2008) Identification of geographically distributed sub-populations of *Leishmania (Leishmania) major* by microsatellite analysis. *BMC Evol Biol* 8: 183.
43. Balloux F, Lehmann L, de Meeus T (2003) The population genetics of clonal and partially clonal diploids. *Genetics* 164: 1635–1644.
44. De Meeus T, Lehmann L, Balloux F (2006) Molecular epidemiology of clonal diploids: A quick overview and a short DIY (do it yourself) notice. *Infect Genet Evol* 6: 163–170.
45. Mark Welch D, Meselson M (2000) Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* 288: 1211–1215.
46. Maynard Smith J, Smith NH, O'Rourke M, Spratt BG (1993) How Clonal are Bacteria? *Proc Natl Acad Sci U S A* 90: 4384–4388.
47. Birky CW Jr (1996) Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics* 144: 427–437.
48. Forche A, Alby K, Schaefer D, Johnson AD, Berman J, et al. (2008) The parasexual cycle in *Candida albicans* provides an alternative pathway to meiosis for the formation of recombinant strains. *PLoS Biol* 6: e110. doi:10.1371/journal.pbio.0060110.
49. Spotorno OA, Córdova L, Solari IA (2008) Differentiation of *Trypanosoma cruzi* I subgroups through characterization of cytochrome b gene sequences. *Infect Genet Evol* 8: 898–900.
50. Westenberger SJ, Sturm NR, Campbell DA (2006) *Trypanosoma cruzi* 5S rRNA arrays define five groups and indicate the geographic origins of an ancestor of the heterozygous hybrids. *Int J Parasitol* 36: 337–346.
51. Brisse S, Verhoef J, Tibayrenc M (2001) Characterisation of large and small subunit rRNA and mini-exon genes further supports the distinction of six *Trypanosoma cruzi* lineages. *Int J Parasitol* 31: 1218–1226.
52. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, eds. *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Totowa (New Jersey): Humana Press.
53. Goudet J (1995) FSTAT Version 1.2: A computer program to calculate F-statistics. *J Heredity* 86: 485–486.
54. Hurlbert S (1971) The non concept of species diversity: A critique and alternative parameters. *Ecology* 52: 577–586.
55. Gutiérrez J, Royo L, Álvarez I, Goyache F (2005) MolKin v2.0: A computer program for genetic analysis of populations using molecular coancestry information. *J Heredity* 96: 718–721.
56. Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1: 47–50.
57. Rice W (1989) Analyzing tables with statistical tests. *Evolution* 43: 223–225.
58. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358–1370.
59. Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Mol Ecol Notes* 1: 101–102.
60. Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza L (1995) MICROSAT—The Microsatellite Distance Program. Stanford: Stanford University Press.
61. Peakall R, Smouse P (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6: 288–295.